

909/897188

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: October 13, 2004, 11:23:49 ; Search time 88.6045 Seconds
(without alignments)
4449.477 Million cell updates/sec

Title: SEQ1-A
Perfect score: 5835
Sequence: 1 MAWKLTPIYLLLSVFEVQ.....DMDYLPVNPQGIINPMLS 1099

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.1

Searched: 2002273 seqs, 358729299 residues

Total number of hits satisfying chosen parameters: 2002273

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A_Geneseq_23Sep04:*
1: Geneseqp1980s:*
2: Geneseqp1990s:*
3: Geneseqp2000s:*
4: Geneseqp2001s:*
5: Geneseqp2002s:*
6: Geneseqp2003as:*
7: Geneseqp2003bs:*
8: Geneseqp2004s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | DB ID | Description |
|------------|--------|-------------|--------|-------|---------------------|
| 1 | 5820.9 | 99.8 | 1299 | 4 | AAM24322 Human EST |
| 2 | 5820.9 | 99.8 | 1404 | 2 | AAR26049 MSF prcu |
| 3 | 5820.9 | 99.8 | 1404 | 4 | AAB29773 Human meg |
| 4 | 5820.9 | 99.8 | 1404 | 4 | AAB60568 Human meg |
| 5 | 5820.9 | 99.8 | 1404 | 8 | ADM98014 Human meg |
| 6 | 5820.9 | 99.8 | 1415 | 4 | AAU32262 Novel hum |
| 7 | 5820.9 | 99.8 | 1404 | 7 | ADK65839 Angiogene |
| 8 | 5808.7 | 96.1 | 1320 | 7 | ADK65819 Angiogene |
| 9 | 5353.7 | 91.8 | 1270 | 8 | ADK67911 Human ext |
| 10 | 5339.6 | 91.5 | 1311 | 8 | ADK67912 Human ext |
| 11 | 2920 | 50.0 | 546 | 4 | ABU53252 Human tes |
| 12 | 2888 | 49.5 | 551 | 4 | ABU53253 Human tes |
| 13 | 2880.1 | 49.4 | 902 | 4 | ABU53254 Human tes |
| 14 | 2757 | 47.2 | 513 | 4 | ABU53254 Human tes |
| 15 | 1954.5 | 33.5 | 452 | 2 | AAO18834 3' cartil |
| 16 | 1409.9 | 24.2 | 538 | 5 | AAO18834 3' cartil |
| 17 | 1401.6 | 24.0 | 292 | 5 | AAU11261 Human HAP |
| 18 | 1354.4 | 23.2 | 5179 | 4 | AAM24516 C899P pre |
| 19 | 1354.4 | 23.2 | 5179 | 6 | ABP55365 Human col |
| 20 | 1354.4 | 23.2 | 5179 | 6 | ABO07258 Human p53 |
| 21 | 1354.4 | 23.2 | 5179 | 7 | ADDA48091 Human Pro |
| 22 | 1354.4 | 23.2 | 5179 | 7 | ADDA4998 Human Pro |
| 23 | 1343.3 | 19.6 | 1684 | 2 | AAW43106 C. thermo |
| 24 | 1131.8 | 19.4 | 8991 | 4 | ABU08487 S. pneumo |
| 25 | 1099.6 | 18.8 | 717 | 4 | ABU53144 Human tes |

ALIGNMENTS

RESULT 1
AAM24322
ID AAM24322 standard; protein; 1299 AA.
XX
AC AAM24322;
XX
DT 12-OCT-2001 (first entry)
XX
DE Human EST encoded protein SEQ ID NO: 1847.
XX
KW Human; sheep; pig; cow; fruit fly; yeast; hamster; macaque; horse;
KW tomato; monkey; dog; sea urchin; expressed sequence tag; EST;
KW diagnostics; forensic test; gene mapping; genetic disorder; biodiversity;
KW Gene therapy; nutrition.
XX
OS Homo sapiens.
XX
PN WO200154477-A2.
XX
PD 02-AUG-2001.
XX
PF 25-JAN-2001; 2001WO-US002687.
XX
PR 25-JAN-2000; 2000US-00491404.
PR 17-JUL-2000; 2000US-00617746.
PR 03-AUG-2000; 2000US-00631451.
PR 15-SEP-2000; 2000US-00663870.
XX
PA (HYSE-) HYSEQ INC.
XX
PI Tang YT, Liu C, Zhou P, Qian XB, Wang Z, Chen R, Asundi V;
PI Cao Y, Drmanac RA, Zhang J, Werhman T;
XX
WPI: 2001-476164/51.
XX
DR N-PSDB; AAH98981.
XX
PT Isolated polypeptide for treatment of diseases, diagnostics, raising
XX antibodies and research use.
XX
PS Claim 20; Page 1198-1201; 12/5pp; English.
XX
CC The present invention provides the protein and coding sequences of novel
XX proteins from a variety of organisms, including human, dog, cat, horse,
XX cow, pig, hamster, monkey, macaque, yeast, bacteria, fruit fly, sea
XX urchin and tomato. These were derived from expressed sequence tags (ESTs)
XX from the organism of interest. They can be used in diagnostics,
XX forensics, gene mapping, identification of mutations, to assess
XX biodiversity and for nutritional purposes. The present sequence is a

| | | | | | | |
|----|--------|------|------|---|----------|--------------------|
| 26 | 1097 | 18.8 | 214 | 4 | ABU53255 | ABU53255 Human tes |
| 27 | 1087.6 | 18.6 | 1795 | 4 | ABB69806 | ABB69806 Drosophil |
| 28 | 1052.6 | 18.0 | 763 | 3 | AAG38942 | AAG38942 Arabidops |
| 29 | 1044.2 | 17.9 | 1049 | 4 | ABB61364 | ABB61364 Drosophil |
| 30 | 1034.5 | 17.7 | 2284 | 4 | ABB71434 | ABB71434 Drosophil |
| 31 | 998.1 | 17.1 | 5703 | 8 | ADL23265 | ADL23265 Human MUC |
| 32 | 996.1 | 17.1 | 2112 | 4 | ABB60403 | ABB60403 Drosophil |
| 33 | 990.5 | 17.0 | 4315 | 5 | ABP43908 | ABP43908 MUC5B par |
| 34 | 987.9 | 16.9 | 770 | 4 | ABU53141 | ABU53141 Human tes |
| 35 | 981 | 16.8 | 778 | 4 | ABU53143 | ABU53143 Human tes |
| 36 | 974.4 | 16.7 | 692 | 4 | ABU53155 | ABU53155 Human tes |
| 37 | 973.4 | 16.7 | 1538 | 7 | AD121202 | AD121202 Novel hum |
| 38 | 972.9 | 16.7 | 1460 | 7 | ADN29110 | ADN29110 Cancer/an |
| 39 | 964.1 | 16.5 | 717 | 4 | ABU53150 | ABU53150 Human tes |
| 40 | 964.1 | 16.5 | 717 | 4 | ABU53149 | ABU53149 Human tes |
| 41 | 964.1 | 16.5 | 717 | 4 | ABU53151 | ABU53151 Human tes |
| 42 | 964.1 | 16.5 | 717 | 4 | ABU53145 | ABU53145 Human tes |
| 43 | 964.1 | 16.5 | 717 | 4 | ABU53148 | ABU53148 Human tes |
| 44 | 964.1 | 16.5 | 717 | 4 | ABU53147 | ABU53147 Human tes |
| 45 | 964.1 | 16.5 | 717 | 4 | ABU53146 | ABU53146 Human tes |

| | | | |
|----|-----|---|-----|
| XX | PS | Claim 1, 2 and 3; Fig 1; 87pp; English. | |
| XX | CC | The sequence given is a full length translation from the megakaryocyte | |
| XX | CC | stimulating factor (MSF) precursor. The sequence covered by exons II, III | |
| XX | CC | and IV encodes megakaryocyte stimulating factor (MSF). This sequence is | |
| XX | CC | modified by the addition of an N-terminal sequence encoding a secretory | |
| XX | CC | leader, an initiating methionine preceding exon II and a terminating | |
| XX | CC | codon following exon IV. The cDNA sequence given contains sequences | |
| XX | CC | derived from human megakaryocyte colony stimulating factor (meg-CSF). | |
| XX | CC | Exon I contains the initiating methionine, and encodes a classical | |
| XX | CC | mammalian protein secretion signal sequence. The sequence encoding the | |
| XX | CC | original meg-CSF includes exons II-IV and is thought to terminate in the | |
| XX | CC | region between amino acid residues 134 - 147. The primary transcript of | |
| XX | CC | this gene may be cleaved in different ways to yield a family of mRNAs | |
| XX | CC | each encoding a different MSF protein. Exons V and VI are thought to be | |
| XX | CC | related to the activity of the factor and are also implicated in the | |
| XX | CC | stability, folding and processing of the molecule. These exons are also | |
| XX | CC | thought to play a role in the observed synergy of MSF with other | |
| XX | CC | cytokines. Exons V - XII are believed to be implicated in the processing | |
| XX | CC | or folding of the appropriate structure of the resulting factor, ie. one | |
| XX | CC | or more of these exons may contain sequences which direct proteolytic | |
| XX | CC | cleavage, adhesion, organisation of the cellular matrix or extracellular | |
| XX | CC | matrix processing. Both naturally occurring and non-naturally occurring | |
| XX | CC | MSF's may be characterised by various combinations of alternatively | |
| XX | CC | spliced exons from this sequence, with the exons spliced together in | |
| XX | CC | differing orders to form different members of the MSF family. (Updated on | |
| XX | CC | 25-MAR-2003 to correct PN field.) | |
| XX | SQ | Sequence 1404 AA; | |
| | | Query Match 99.8%; Score 5820.9; DB 2; Length 1404; | |
| | | Best Local Similarity 96.4%; Pred. No. 6.1e-157; | |
| | | Matches 1099; Conservative 0; Mismatches 0; Indels 41; Gaps 1; | |
| QY | 1 | MAWKTLPIYLLLLSVFIQVSSQ----- | 25 |
| DB | 1 | MAWKTLPIYLLLLSVFIQVSSQDLSCACGCGEGYSDATCNCYNCQHMECCPDF | 60 |
| QY | 26 | -----ELSCXGRCFESFERGECDCDAOCKYDKCCPDYEFCAEVHNPTSPSSKKAP | 79 |
| DB | 61 | KVCTAELSCXGRCFESFERGECDCDAOCKYDKCCPDYEFCAEVHNPTSPSSKKAP | 120 |
| QY | 80 | PPSGASQTIKSTIKRSPKPNKKTKVIESBEITEHSHSVSNQESSSSSSSSSTIW | 139 |
| DB | 121 | PPSGASQTIKSTIKRSPKPNKKTKVIESBEITEHSHSVSNQESSSSSSSSSTIW | 180 |
| QY | 140 | KIKSSKNSAANRELQKLVKDKNKKTKKTPKPPVVDGAGSLDNGDFKVTTPDST | 199 |
| DB | 181 | KIKSSKNSAANRELQKLVKDKNKKTKKTPKPPVVDGAGSLDNGDFKVTTPDST | 240 |
| QY | 200 | TOHNKYSTGPKITAKPINRPSLPNSDTSKETSITVNKETTIVETKTTNKQSTDG | 259 |
| DB | 241 | TOHNKYSTGPKITAKPINRPSLPNSDTSKETSITVNKETTIVETKTTNKQSTDG | 300 |
| QY | 260 | KEKTTSAKETQSIKTSKADLAPTSKVLAKPTPKAETTTKGPALTTPKPTTTPKEPAS | 319 |
| DB | 301 | KEKTTSAKETQSIKTSKADLAPTSKVLAKPTPKAETTTKGPALTTPKPTTTPKEPAS | 360 |
| QY | 320 | TTPKPTTTPKEPASPTTTPKEPASPTTTPKEPASPTTTPKEPASPTTTPKEPAS | 379 |
| DB | 361 | TTPKPTTTPKEPASPTTTPKEPASPTTTPKEPASPTTTPKEPASPTTTPKEPAS | 420 |
| QY | 380 | APTTTKSAPTTTPKEPASPTTTPKEPASPTTTPKEPASPTTTPKEPASPTTTPK | 439 |
| DB | 421 | APTTTKSAPTTTPKEPASPTTTPKEPASPTTTPKEPASPTTTPKEPASPTTTPK | 480 |
| QY | 440 | EPAPTAKKAPPTTPKEPASPTTTPKEPASPTTTPKEPASPTTTPKEPASPTTTPK | 499 |
| DB | 481 | EPAPTAKKAPPTTPKEPASPTTTPKEPASPTTTPKEPASPTTTPKEPASPTTTPK | 540 |
| QY | 500 | TTKSAPTTTPKEPASPTTTPKEPASPTTTPKEPASPTTTPKEPASPTTTPKEPAS | 559 |

| | | | |
|----|----------|---|------|
| DB | 541 | TTKSAPTTTPKEPASPTTTPKEPASPTTTPKEPASPTTTPKEPASPTTTPKEPAS | 600 |
| QY | 560 | APTAPKEPASPTTTPKEPASPTTTPKEPASPTTTPKEPASPTTTPKEPASPTTTPK | 619 |
| DB | 601 | APTAPKEPASPTTTPKEPASPTTTPKEPASPTTTPKEPASPTTTPKEPASPTTTPK | 660 |
| QY | 620 | PPEAPPTTTPKAAAPNTPKPEPASPTTTPKEPASPTTTPKEPASPTTTPKEPAS | 679 |
| DB | 661 | PPEAPPTTTPKAAAPNTPKPEPASPTTTPKEPASPTTTPKEPASPTTTPKEPAS | 720 |
| QY | 680 | APTTPKKKAPKELAPTTTKEPTSTTSDDKAPTTTSGTAPTTTTPKEPASPTTTPK | 739 |
| DB | 721 | APTTPKKKAPKELAPTTTKEPTSTTSDDKAPTTTSGTAPTTTTPKEPASPTTTPK | 780 |
| QY | 740 | TAPTTLKEPASPTTTPKKKAPKELAPTTTKEPTSTTSDDKAPTTTSGTAPTTTTPK | 799 |
| DB | 781 | TAPTTLKEPASPTTTPKKKAPKELAPTTTKEPTSTTSDDKAPTTTSGTAPTTTTPK | 840 |
| QY | 800 | KPAPTTPETPPPTTSEVSTPTTKEPTTTHKSPDESTPELSAETPKALENSPKGVP | 859 |
| DB | 841 | KPAPTTPETPPPTTSEVSTPTTKEPTTTHKSPDESTPELSAETPKALENSPKGVP | 900 |
| QY | 860 | TKTPAATKPEMTTAKDKITERDLRTTPTTAAAPKMTKETATTTTEKTTESKITAT | 919 |
| DB | 901 | TKTPAATKPEMTTAKDKITERDLRTTPTTAAAPKMTKETATTTTEKTTESKITAT | 960 |
| QY | 920 | TSTTTQDTTTPFKITLTKTTTLPKVTITTKITTTIMMKPESTAKPKRATNSKATTP | 979 |
| DB | 961 | TSTTTQDTTTPFKITLTKTTTLPKVTITTKITTTIMMKPESTAKPKRATNSKATTP | 1020 |
| QY | 980 | PQKPTKAPKPTSTKPKMTMPVRKKTTPTRKMTSTMPELNPTSRIAEAMLOTTTRPN | 1039 |
| DB | 1021 | PQKPTKAPKPTSTKPKMTMPVRKKTTPTRKMTSTMPELNPTSRIAEAMLOTTTRPN | 1080 |
| QY | 1040 | QTNSKLVENVNPKSEAGGAEGETPHMLLRPHVFMPEVTPDMYLPVNPQGIINPMLS | 1099 |
| DB | 1081 | QTNSKLVENVNPKSEAGGAEGETPHMLLRPHVFMPEVTPDMYLPVNPQGIINPMLS | 1140 |
| | RESULT 3 | | |
| | AAB29773 | | |
| ID | AAB29773 | standard; protein; 1404 AA. | |
| XX | AC | AAB29773; | |
| XX | DT | 28-FEB-2001 (first entry) | |
| XX | DE | Human megakaryocyte stimulating factor (MSF), SEQ ID NO:1. | |
| XX | KW | Human MSF; megakaryocyte stimulating factor; tribonectin; | |
| XX | KW | alternative splicing; joint boundary lubricant; O-linked oligosaccharide; | |
| XX | KW | osteoarthritis; tribosupplementation; tissue adhesion inhibition; | |
| XX | OS | friction coefficient reduction; gene therapy; antiarthritic; osteopathic. | |
| XX | OS | Homo sapiens. | |
| XX | FN | WO200064930-A2. | |
| XX | PD | 02-NOV-2000. | |
| XX | PF | 24-APR-2000; 2000WO-US010953. | |
| XX | PR | 23-APR-1999; 99US-00298970. | |
| XX | XX | (RHOD-) RHODE ISLAND HOSPITAL LIFESPAN PARTNER. | |
| PI | PI | Jay GD; | |
| XX | XX | WPI; 2001-024673/03. | |
| DR | DR | N-PSDB; AAC81498. | |
| XX | PT | Novel tribonectin polypeptide useful as lubricant for treating | |
| PT | PT | osteoarthritis, comprises O-linked lubricating moiety. | |

XX PS Claim 3; Page 7; 47pp; English.

XX CC The invention relates to a human tribonection which is a product of

CC alternative splicing of the human MSF (megakaryocyte stimulating factor)

CC gene. The tribonection has at least one O-linked oligosaccharide

CC lubricating moiety and has a polypeptide sequence comprising 1-76 repeats

CC of a motif having at least 50% identity to the sequence KQPAPTT

CC (AAB29774). The invention also relates to a nucleic acid encoding a human

CC MSF-derived tribonection; a biocompatible composition comprising a human

CC tribonection for inhibiting tissue adhesion formation; and a method of

CC diagnosing osteoarthritis or a predisposition to osteoarthritis by

CC measuring the amount of MSF or its fragment in a biological sample of a

CC mammal, wherein an increased amount of MSF compared to a control

CC indicates the presence of or predisposition to developing osteoarthritis.

CC The tribonection and DNA encoding it are useful in the treatment of

CC osteoarthritis, where they may be used for lubricating mammalian joints,

CC such as articulating joints of humans, dogs or horses. The tribonection,

CC when formulated as a membrane, foam, gel or fibre, is useful for

CC inhibiting adhesion between two surfaces such as the injured tissues of a

CC mammal, where the injury is caused by a surgical insertion or trauma, or

CC an artificial device e.g., an orthopaedic implant. In particular, one of

CC the surfaces is pericardial tissue. DNA encoding a tribonection may be

CC used in gene therapy. The present sequence represents human MSF

XX SQ Sequence 1404 AA;

Query Match 99.8%; Score 5820.9; DB 4; Length 1404;

Best Local Similarity 96.4%; Pred. No. 6.1e-157;

Matches 1099; Conservative 0; Mismatches 0; Indels 41; Gaps 1;

QY 1 MAWKTLPIYLLLSVFIQVSSQ----- 25

DB 1 MAWKTLPIYLLLSVFIQVSSQDLSACRGCGRGYSRATCNCNDYHCQHYMECCPDF 60

QY 26 -----ELSCGKGFESPERGECDCDAQCKYKCCPDYBSFCAEVHNPTSPSSKAP 79

DB 61 KRVCYTAELSCGKGFESPERGECDCDAQCKYKCCPDYBSFCAEVHNPTSPSSKAP 120

QY 80 PPSGASQIKSTTKSPDPNKKTKKVIIEBETEEHSVSENOESSSSSSSSSSSTIW 139

DB 121 PPSGASQIKSTTKSPDPNKKTKKVIIEBETEEHSVSENOESSSSSSSSSSSTIW 180

QY 140 KIKSSKNSAANRELQKLVKDNKNRTKKXPTPKPVVDAGSLDNGPKVTTPTST 199

DB 161 KIKSSKNSAANRELQKLVKDNKNRTKKXPTPKPVVDAGSLDNGPKVTTPTST 240

QY 200 TQHNKVSTSPKITTAKPINRPSLPNDSKTSKSLTVNKETTVETKETTINKQSTDG 259

DB 241 TQHNKVSTSPKITTAKPINRPSLPNDSKTSKSLTVNKETTVETKETTINKQSTDG 300

QY 260 KEKTTISAKETOSIEKTSKADLAPTSKVLAKPTPKAETTTKGPALTTKPEPTTPKBPAS 319

DB 301 KEKTTISAKETOSIEKTSKADLAPTSKVLAKPTPKAETTTKGPALTTKPEPTTPKBPAS 360

QY 320 TTPKEPTTTIKSAPTTKPEAPTTTKSAPTTKPEAPTTTKPEAPTTTKPEAPTTTKPE 379

DB 361 TTPKEPTTTIKSAPTTKPEAPTTTKSAPTTKPEAPTTTKPEAPTTTKPEAPTTTKPE 420

QY 380 APTTKSAPTTKPEAPTTTKKRAPTTKPEAPTTKPEPTTPKPEAPTTKPEAPTTTKP 439

DB 421 APTTKSAPTTKPEAPTTTKKRAPTTKPEAPTTKPEPTTPKPEAPTTKPEAPTTTKP 480

QY 440 EPAPTAPKPAPTTKPEAPTTTKPEAPTTTKPEAPTTTKPEAPTTTKSAPTTKPEAPT 499

DB 481 EPAPTAPKPAPTTKPEAPTTTKPEAPTTTKPEAPTTTKPEAPTTTKSAPTTKPEAPT 540

QY 500 TTKSAPTTKPEAPTTTKPEAPTTTKPEAPTTTKKPAPTTKPEAPTTKPEAPTTTKK 559

DB 541 TTKSAPTTKPEAPTTTKPEAPTTTKPEAPTTTKKPAPTTKPEAPTTKPEAPTTTKK 600

QY 560 APTAPKPEAPTTTKPEAPTTPKKLTPTTPEKLAPTTKPEAPTTPEELAPTTPEEPTPT 619

DB 601 APTAPKPEAPTTTKPEAPTTPKKLTPTTPEKLAPTTKPEAPTTPEELAPTTPEEPTPT 660

QY 620 PEEPAPTPKAAAPNTPKPEAPTTPKPEAPTTPKPEAPTTPKETAPTTPKGTAPTTLKEP 679

DB 661 PEEPAPTPKAAAPNTPKPEAPTTPKPEAPTTPKPEAPTTPKETAPTTPKGTAPTTLKEP 720

QY 680 APTTPKPAKELAPTTTKETSTSDKPAPTTKGTAPTTPKGTAPTTPKPEAPTTPKG 739

DB 721 APTTPKPAKELAPTTTKETSTSDKPAPTTKGTAPTTPKGTAPTTPKPEAPTTPKG 780

QY 740 TAPTTLKEAPTTPKPAKELAPTTTKGTSTSDKPAPTTKETAPTTPKPEAPTTPK 799

DB 781 TAPTTLKEAPTTPKPAKELAPTTTKGTSTSDKPAPTTKETAPTTPKPEAPTTPK 840

QY 800 KPAPTTPETPPPTTSEVSTPTTKETTHKSDPSTPSELSAAPTTPKALENSPKPEGVT 859

DB 841 KPAPTTPETPPPTTSEVSTPTTKETTHKSDPSTPSELSAAPTTPKALENSPKPEGVT 900

QY 860 TKTPAATKPEMTTAKDKITERDLRTPETTTTAAAPKMTKETATTTTETTESKITATTQV 919

DB 901 TKTPAATKPEMTTAKDKITERDLRTPETTTTAAAPKMTKETATTTTETTESKITATTQV 960

QY 920 TSTTTQDTPPFKLTTLTKTTLAPKVTTKTITTTIMNKPBETAKPKDRATNSKATTPK 979

DB 961 TSTTTQDTPPFKLTTLTKTTLAPKVTTKTITTTIMNKPBETAKPKDRATNSKATTPK 1020

QY 980 PQKPTKAPKPTSTKPKTMPRVKPKTTPKMTSTMPKLTSTMPKLTSTMPKLTSTMPKLTSTMPK 1039

DB 1021 PQKPTKAPKPTSTKPKTMPRVKPKTTPKMTSTMPKLTSTMPKLTSTMPKLTSTMPKLTSTMPK 1080

QY 1040 QTPNSKLVEVNPKEGDAGAGETPHMLLRPHVFPVETPDMDYLPRVFNQGIINPMLS 1099

DB 1081 QTPNSKLVEVNPKEGDAGAGETPHMLLRPHVFPVETPDMDYLPRVFNQGIINPMLS 1140

RESULT 4

ID AAB60568 standard; protein; 1404 AA.

XX AC AAB60568;

XX XX

DT 27-APR-2001 (first entry)

XX XX

DE Human megakaryocyte stimulating factor (MSF, CACP).

XX XX

KW Human; CACP protein; camptodactylly-arthropathy-coxa vara-pericarditis;

KW MSF; megakaryocyte stimulating factor; synovial lubricant;

KW chromosome 1q25-31; osteoarthritis; joint lubrication; osteopathic;

XX antiarthritic.

XX OS Homo sapiens.

XX XX

PN WO200107068-A1.

XX XX

PD 01-FEB-2001.

XX XX

PF 21-JUL-2000; 2000WO-US020002.

XX XX

PR 23-JUL-1999; 99US-0145328P.

PR 19-JUL-2000; 2000US-00145328.

XX XX

PA (UYCA-) UNIV CASE WESTERN RESERVE.

XX XX

PI Warman ML;

XX XX

DR WPI; 2001-192721/18.

XX XX

PT New composition comprising the camptodactylly-arthropathy-coxa vara-

PT pericarditis protein in combination with an anesthetic, useful for

PT treating osteoarthritis, or as lubricants of tissue and joints.

XX XX

PS Example 1; Page; 34pp; English.

XX XX

CC The invention relates to a method of treating osteoarthritis via the
 CC administration of a composition comprising the camptodactyl-arthropathy-
 CC coxa vara-pericarditis (CACP) protein, or portions of the CACP protein.
 CC The composition may further comprise a local anaesthetic. The composition
 CC of the invention may be administered via intra-articular or intravenous
 CC injection. The human CACP protein is identified in the invention as being
 CC megakaryocyte stimulating factor (MSF). The gene encoding CACP protein
 CC (MSF) is located on chromosome 1q25-31, and mutations in this gene are
 CC responsible for the heritable disorder camptodactyl-arthropathy-coxa
 CC vara-pericarditis, in which patients have synovial hyperplasia without
 CC evidence of inflammation. CACP protein (MSF) acts as a synovium
 CC lubricant, and can be used to lubricate tissue and joints in the
 CC treatment of osteoarthritis. The composition may be applied to reduce the
 CC symptoms of osteoarthritis (e.g., joint pain, loss of range of movement
 CC or joint damage). The present sequence represents human megakaryocyte
 CC stimulating factor (MSF, CACP protein). Note: This sequence is not given
 CC in its entirety in figure 4 of the specification, although a GenBank
 CC accession number was given. This sequence was therefore obtained from
 CC GenBank (U70316)
 CC XX
 CC SQ Sequence 1404 AA;

Query Match 99.8%; Score 5820.9; DB 4; Length 1404;
 Best Local Similarity 96.4%; Pred. No. 6.1e-157;
 Matches 1099; Conservative 0; Mismatches 0; Indels 41; Gaps 1;

QY 1 MAWKTLPIYLLLSVFIQVSSQ----- 25
 DB 1 MAWKTLPIYLLLSVFIQVSSQ----- 25
 QY 26 -----ELCKGCRGPFERECDCDAOCKYDKCCPDYEFCAEVHNPSPSPSKAP 79
 DB 61 KRVCTAELCKGRCFSEFGRRECDCDAOCKYDKCCPDYEFCAEVHNPSPSPSKAP 120
 QY 80 PPSGASQTIKSTTKRSPKPNKKTKVIESEIIEHVSQNSQSSSSSSSSSSSTIW 139
 DB 121 PPSGASQTIKSTTKRSPKPNKKTKVIESEIIEHVSQNSQSSSSSSSSSSSTIW 180
 QY 140 KIKSSNSAANRELQKLVKQKRTKKTPKPPVDEAGSLDNGDFKVTTPDST 199
 DB 181 KIKSSNSAANRELQKLVKQKRTKKTPKPPVDEAGSLDNGDFKVTTPDST 240
 QY 200 TQNKVSTSPKITTAKPINRPSLPNSDTSKETSITVKNKETTVEKTTTNKQSTSDG 259
 DB 241 TQNKVSTSPKITTAKPINRPSLPNSDTSKETSITVKNKETTVEKTTTNKQSTSDG 300
 QY 260 KEKTSAGKTSQIEKTSAXDLAFTSKVLAKPTPKAETTTKGPALITPKPPTTPKEPAS 319
 DB 301 KEKTSAGKTSQIEKTSAXDLAFTSKVLAKPTPKAETTTKGPALITPKPPTTPKEPAS 360
 QY 320 TTPKEPTPTIKSAPTTKPEPATTTKSAPTTKPEPATTTPKEPATTTPKEPATTTPKEP 379
 DB 361 TTPKEPTPTIKSAPTTKPEPATTTKSAPTTKPEPATTTPKEPATTTPKEPATTTPKEP 420
 QY 380 APTTTKSAPTTTPKEPATTTPKEPATTTPKEPATTTPKEPATTTPKEPATTTPKEP 439
 DB 421 APTTTKSAPTTTPKEPATTTPKEPATTTPKEPATTTPKEPATTTPKEPATTTPKEP 480
 QY 440 EPAPTAKPAPTTPKEPATTTPKEPATTTPKEPATTTPKEPATTTPKEPATTTPKEP 499
 DB 481 EPAPTAKPAPTTPKEPATTTPKEPATTTPKEPATTTPKEPATTTPKEPATTTPKEP 540
 QY 500 TTKSAPTTTPKEPSTTTKPEPATTTPKEPATTTPKEPATTTPKEPATTTPKEPATTTPKEP 559
 DB 541 TTKSAPTTTPKEPSTTTKPEPATTTPKEPATTTPKEPATTTPKEPATTTPKEPATTTPKEP 600
 QY 560 APTAPKEPAPTTPKEPAPTTPKLTTPPTPEKLAPTTPKEPAPTTPBELAPTTPPEPTPT 619
 DB 601 APTAPKEPAPTTPKEPAPTTPKLTTPPTPEKLAPTTPKEPAPTTPBELAPTTPPEPTPT 660
 QY 620 PEEAPPTPKAAANTPKAPPTTPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKEP 679
 DB 661 PEEAPPTPKAAANTPKAPPTTPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKEP 720

QY 680 APTTPKKPAPKELAPTTTKEPTSTTSKDPAPTPPKGTAPTTPKPAPTTPKPAPTTPKG 739
 DB 721 APTTPKKPAPKELAPTTTKEPTSTTSKDPAPTPPKGTAPTTPKPAPTTPKPAPTTPKG 780
 QY 740 TAPTTLKEPAPTTPKPAPKELAPTTTKEPTSTTSKDPAPTTPKGTAPTTPKPAPTTPK 799
 DB 781 TAPTTLKEPAPTTPKPAPKELAPTTTKEPTSTTSKDPAPTTPKGTAPTTPKPAPTTPK 840
 QY 800 KPAPTTPPETPTTSEVSTPTTKEPTTHKSPDESTPELSABPTPKALENSPKPBGVPT 859
 DB 841 KPAPTTPPETPTTSEVSTPTTKEPTTHKSPDESTPELSABPTPKALENSPKPBGVPT 900
 QY 860 TKPAAKTPMTTAKDKTTERDLRTPTTAAAPKMTKETATTTKTTESKITATTTOV 919
 DB 901 TKPAAKTPMTTAKDKTTERDLRTPTTAAAPKMTKETATTTKTTESKITATTTOV 960
 QY 920 TSITTTQDTTTPFKITTLKTTTLAPKVTITTKTITTEIMNKPEETAKPKDASATNSKATPK 979
 DB 961 TSITTTQDTTTPFKITTLKTTTLAPKVTITTKTITTEIMNKPEETAKPKDASATNSKATPK 1020
 QY 980 PQKPTKAPKPTSTTKKKTMTPRVRKPKTTTTPRMTSTMPELNPTSRIAEAMLOTTTRN 1039
 DB 1021 PQKPTKAPKPTSTTKKKTMTPRVRKPKTTTTPRMTSTMPELNPTSRIAEAMLOTTTRN 1080
 QY 1040 QTPNSKLVEVNPXSSEDAGGAEGETPHMLLRPHVFMPEVTPDMDYLPVNPQGIINPMLS 1099
 DB 1081 QTPNSKLVEVNPXSSEDAGGAEGETPHMLLRPHVFMPEVTPDMDYLPVNPQGIINPMLS 1140
 RESULT 5
 ADM98014
 ID ADM98014 standard; pro-ein; 1404 AA.
 XX
 AC ADM98014;
 XX
 DT 01-JUL-2004 (first entry)
 XX
 DE Human megakaryocyte stimulating factor (MSF).
 XX
 KW lubricating polypeptide; O-linked oligosaccharide; joint lubrication;
 KW CAP; camptodactyl-arthropathy pericarditis; osteoarthritis; human;
 KW megakaryocyte stimulating factor; MSF.
 XX
 OS Homo sapiens.
 XX
 EN US2004072741-A1.
 XX
 PD 15-APR-2004.
 XX
 PF 02-JUL-2001; 2001US-00897188.
 XX
 PR 23-APR-1999; 99US-00298970.
 PR 24-APR-2000; 2000US-00556246.
 XX
 PA (JAYG/) JAY G D.
 XX
 PI Jay GD;
 XX
 XX WPI: 2004-373948/35.
 DR N-PSDB; ADM98015.
 XX
 PT New tribonectin polypeptides and polynucleotides for lubricating joints
 PT or other tissues to prevent or treat Camptodactyl-arthropathy-
 PT pericarditis syndrome or osteoarthritis.
 XX
 PS Claim 1; SEQ ID NO 1; 34pp; English.
 XX
 CC The invention relates to a lubricating polypeptide and at least one O-
 CC linked oligosaccharide. The composition and methods are useful for
 CC lubricating joints or other tissues to prevent or treat camptodactyl-
 CC arthropathy pericarditis (CAP) or osteoarthritis in mammals. The present
 CC sequence represents the amino acid sequence of the human megakaryocyte

Qv 920 TSTTTODTTPFKITTLKTTTLAPKVTTTCKT TTTTETMKNKPEETAKPKDPRATNSKATTPK 929

Qy

1 MAY 1997

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
84

Qy 1 MAWKTLPIYLLLSVFVIQQVSSQ----- 25

QY 140 KIKSSKNSAANRELOKLVKDKNKKQRTKKKTPPKPPVVDEAGSGLDNGDFKVTTPDTST 199
Db 181 KIKSSKNSAANRELOKLVKDKNKKQRTKKKTPPKPPVVDEAGSGLDNGDFKVTTPDTST 240
QY 200 TQHNKYSTSPKITTAKPKNRPSLPNSDTSKETSIVNKKETVETKETTINKSTSDG 259
Db 241 TQHNKYSTSPKITTAKPKNRPSLPNSDTSKETSIVNKKETVETKETTINKSTSDG 300
QY 260 KETTSKETSQTSKETSADKAPTSKVLAKPTKPAETTTKGPALTPKPEPTTTKPEPAS 319
Db 301 KETTSKETSQTSKETSADKAPTSKVLAKPTKPAETTTKGPALTPKPEPTTTKPEPAS 360
QY 320 TTPKEPTPTTKSAPTTKPEAPTTTKSAPTTKPEAPTTTKPEAPTTTKPEAPTTTKPE 379
Db 361 TTPKEPTPTTKSAPTTKPEAPTTTKSAPTTKPEAPTTTKPEAPTTTKPEAPTTTKPE 420
QY 380 APTTTKSAPTTKPEAPTTTKKPAPTTKPEAPTTTKPEPTTTKPEAPTTTKPEAPTTTK 439
Db 421 APTTTKSAPTTKPEAPTTTKKPAPTTKPEAPTTTKPEPTTTKPEAPTTTKPEAPTTTK 480
QY 440 EPAPTAPKPAPTTKPEAPTTTKKPAPTTKPEAPTTTKPEPTTTKPEAPTTTKSAPTTK 499
Db 481 EPAPTAPKPAPTTKPEAPTTTKKPAPTTKPEAPTTTKPEPTTTKPEAPTTTKSAPTTK 540
QY 500 TTKSAPTTTKPEPTTKPEAPTTTKKPAPTTKPEAPTTTKPEPTTKKPAPTTKKPP 559
Db 541 TTKSAPTTTKPEPTTKPEAPTTTKKPAPTTKPEAPTTTKPEPTTKKPAPTTKKPP 600
QY 560 APTAPKPAPTTKKPAPTTKKPAPTTKKPAPTTKKPAPTTKKPAPTTKKPAPTTKK 619
Db 601 APTAPKPAPTTKKPAPTTKKPAPTTKKPAPTTKKPAPTTKKPAPTTKKPAPTTKK 660
QY 620 PEPAPTTPKAAAPTTPKPAPTTKPEAPTTTKPEAPTTTKPEAPTTTKPEAPTTTKPE 679
Db 661 PEPAPTTPKAAAPTTPKPAPTTKPEAPTTTKPEAPTTTKPEAPTTTKPEAPTTTKPE 720
QY 680 APTTPKPAKPAELAPTTTKPEPTSTSDKPAPTTKGTAAPTTPKPAPTTKPEAPTTPK 739
Db 721 APTTPKPAKPAELAPTTTKPEPTSTSDKPAPTTKGTAAPTTPKPAPTTKPEAPTTPK 780
QY 740 TAPTTLKPAPTTKKPAKPAELAPTTTKGPTSTSDKPAPTTKGTAAPTTPKPEAPTTPK 799
Db 781 TAPTTLKPAPTTKKPAKPAELAPTTTKGPTSTSDKPAPTTKGTAAPTTPKPEAPTTPK 840
QY 800 KPAPTTPPTTSEVSTPTTKPEPTTIHKSPDESTPELSAAPTPKALENSKPEGVPT 859
Db 841 KPAPTTPPTTSEVSTPTTKPEPTTIHKSPDESTPELSAAPTPKALENSKPEGVPT 900
QY 860 TKTPAATPEMTTAKDKTTERDLATTTTETTAAPKMTKETATTEKTESKITATTTOV 919
Db 901 TKTPAATPEMTTAKDKTTERDLATTTTETTAAPKMTKETATTEKTESKITATTTOV 960
QY 920 TSTTTQDTPPKITTLKTTTLAPKVTITTKKITTTEIMNKPEETAKPKDRATNSKATTPK 979
Db 961 TSTTTQDTPPKITTLKTTTLAPKVTITTKKITTTEIMNKPEETAKPKDRATNSKATTPK 1020
QY 980 POKPTKAPKPTSTTKPKTMRVRKPTTTPRKTSTMPNLTSPRIAEAMLTTTRPN 1039
Db 1021 POKPTKAPKPTSTTKPKTMRVRKPTTTPRKTSTMPNLTSPRIAEAMLTTTRPN 1080
QY 1040 QTPNSKLVNPNKSDAGAGETPHMLLRHVFWPEVTPDMYLPVFNQIINPMLS 1099
Db 1081 QTPNSKLVNPNKSDAGAGETPHMLLRHVFWPEVTPDMYLPVFNQIINPMLS 1140

RESULT 8
ID ADK65819 standard; protein; 1320 AA.
XX ADK65819;
AC ADK65819;
XX 06-MAY-2004 (first entry)
DT XX

DE Angiogenesis-differentially expressed protein ANH0316.
XX cystostatic; cardiant; vasotropic; antiarteriosclerotic;
KW angiogenesis inhibitor; angiogenesis stimulator; angiogenic index;
KW gene expression; cancer; coronary artery disease; myocardial ischemia;
KW coronary arteriosclerosis; forensic medicine.
OS Homo sapiens.
XX WO2003066831-A2.
XX 14-AUG-2003.
XX 07-FEB-2003; 2003WO-US003848.
XX 07-FEB-2002; 2002US-00067482.
PR 10-JUN-2002; 2002US-00164595.
PR 16-AUG-2002; 2002US-0403649P.
PR 03-JAN-2003; 2003US-0437746P.
XX (ORIG-) ORIGENE TECHNOLOGIES INC.
XX Sun Z, Li X, Kovacs KF, Fan W, Jay G;
XX WPI; 2003-731502/69.
XX N-PSDB; ADK65818.
XX Determining the angiogenic index of a tissue or cell sample using
PT expression levels of differentially expressed genes, useful for
PT diagnosing or treating cancer, coronary artery disease, myocardial
PT ischemia and/or arteriosclerosis.
XX Claim 23; SEQ ID NO 58; 296pp; English.
XX The invention relates to a method of determining the angiogenic index of
CC a tissue or cell sample comprising assessing, in a sample, the expression
CC levels of one or more differentially-expressed gene from any of 34 DNA
CC sequences, given in the specification, where the levels are indicative of
CC the angiogenic index. The methods and compositions of the present
CC invention are useful for diagnosing, preventing and/or treating cancer,
CC coronary artery disease, myocardial ischemia or coronary
CC arteriosclerosis. They can also be used in research, drug discovery and
CC forensic medicine involving angiogenesis. This sequence corresponds to
CC one of the differentially expressed proteins of the invention.
XX SQ Sequence 1320 AA;
Query Match 96.1%; Score 5608.7; DB 7; Length 1320;
Best Local Similarity 95.9%; Pred. No. 5.9e-151;
Matches 1054; Conservative 0; Mismatches 2; Indels 43; Gaps 1;
QY 1 MAWKTLPIVLLLLSVFVIQVSSOELSCGRCFSGFERGECDDAOCKYDKCCPDYE 60
Db 1 MAWKTLPIVLLLLSVFVIQVSSOELSCGRCFSGFERGECDDAOCKYDKCCPDYE 60
QY 61 SFCAEVHNPSTSPSKKAPPPSGASQTIKTTKRSKPPNKKTKVIESEBITEHSVS 120
Db 61 SFCAEVHNPSTSPSKKAPPPSGASQTIKTTKRSKPPNKKTKVIESEBITE 115
QY 121 ENQESSSSSSSSSSSTTIWKIKSSKNSAANRELOKLVKDKNKKQRTKKKTPPKPVVDE 180
Db 116 -----VKDNKQRTKKKTPPKPVVDE 137
QY 181 AGSGLDNGDFKVTTPDTSTTQHNKYSTSPKITTAKPKNRPSLPNSDTSKETSIVNKK 240
Db 138 AGSGLDNGDFKVTTPDTSTTQHNKYSTSPKITTAKPKNRPSLPNSDTSKETSIVNKK 197
QY 241 TTVETKETTINKSTSDGKETSADKAPTSKVLAKPTKPAETTTKGPALTPKPEPTTK 300
Db 198 TTVETKETTINKSTSDGKETSADKAPTSKVLAKPTKPAETTTKGPALTPKPEPTTK 257
QY 301 PALTTPKPEPTTKPEAPTTTKKPAPTTKPEAPTTTKPEAPTTTKSAPTTTKPEAPTT 360

Db 258 PALTTPKBPTTTPKEPASTTTPKEPTTTIKSAPTTTPKEPAPTTTKSAPTTTPKEPAPTTT 317
Qy 361 KEPAATTPKEPAPTTTKSAPTTTPKEPAPTTTKSAPTTTPKEPAPTTTKSAPTTTPKEPAPTT 420
Db 318 KEPAATTPKEPAPTTTKSAPTTTPKEPAPTTTKSAPTTTPKEPAPTTTKSAPTTTPKEPAPTT 377
Qy 421 TTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTT 480
Db 378 TTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTT 437
Qy 481 PAPTTTKSAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTT 540
Db 438 PAPTTTKSAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTT 497
Qy 541 KEPAATTPKEPAPTTTKSAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTT 600
Db 498 KEPAATTPKEPAPTTTKSAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTT 557
Qy 601 PTPPELAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTP 660
Db 558 PTPPELAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTP 617
Qy 661 KETAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKE 720
Db 618 KETAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKE 677
Qy 721 PKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKE 780
Db 678 PKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKE 737
Qy 781 TPKEATTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKE 840
Db 738 TPKEATTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKE 797
Qy 841 ABPTPKALENSKPEGVPTTKPAATKPEMTTAKDKTERDLRTPTTTTAAAPQWTKET 900
Db 798 ABPTPKALENSKPEGVPTTKPAATKPEMTTAKDKTERDLRTPTTTTAAAPQWTKET 857
Qy 901 ATTEKTESKATTTTQVTSNTTQDTPPEKLTTLTKTLTLPKVTTKTKTLTTEIMNKP 960
Db 858 ATTEKTESKATTTTQVTSNTTQDTPPEKLTTLTKTLTLPKVTTKTKTLTTEIMNKP 917
Qy 961 EETAKPKDRAATNSKATTPKQPKPTKAPKPTSTKPKTMVPRVRKPKTPTTPRKMSTMP 1020
Db 918 EETAKPKDRAATNSKATTPKQPKPTKAPKPTSTKPKTMVPRVRKPKTPTTPRKMSTMP 977
Qy 1021 LNPTSRIBAMLOTTTRPNQTSKLVNPNKSEBAGGAGETPHMLLRPHVFMPEVTPD 1080
Db 978 LNPTSRIBAMLOTTTRPNQTSKLVNPNKSEBAGGAGETPHMLLRPHVFMPEVTPD 1037
Qy 1081 MDYLPVNPQGIINPMLS 1099
Db 1038 MDYLPVNPQGIINPMLS 1056

Key Peptide Location/Qualifiers
1. .29 /label= Signal_peptide
/note= "Spans residues 1 to 18, 20, 21, 24, 27 or 29 according to identification method"

WC2004013292-A2.
12-FEB-2004.
30-JUL-2003; 2003WO-US024084.
02-AUG-2002; 2002US-0400810P.
19-SEP-2002; 2002US-0412197P.
04-OCT-2002; 2002US-0416004P.
08-NOV-2002; 2002US-0424862P.
(INCY-) INCYTE CORP.

Elliott VS, Khare R, Tran UK, Swarnakar A, Marquis JP, Richardson TW, Emerling BM, Lindquist EA, Chawla NK, Ramkumar J, Lee SY;
WPI; 2004-157116/15.
N-PSDB; ADK67916.

New extracellular messengers and nucleic acids, useful for diagnosing, treating or preventing e.g. allergies, anemia, Crohn's disease, diabetes mellitus, myasthenia gravis, osteoporosis, psoriasis, cancers or autoimmune thyroiditis.

Claim 59; SEQ ID NO 4; 165pp; English.

The present sequence is that of novel human extracellular messenger (EXMES) incyte ID NO: 7513017CD1 polypeptide. The protein shows homology to human megakaryocyte stimulating factor. The invention provides EXMES polynucleotides and polypeptides, as well as expression vectors, host cells, antibodies, agonists and antagonists, and methods for diagnosing, treating or preventing disorders associated with aberrant expression of EXMES especially autoimmune and inflammatory disorders, cell proliferative disorders and endocrine disorders, e.g. adult respiratory distress syndrome, allergy, anaemia, erythroblastosis foetalis, Crohn's disease, diabetes mellitus, myasthenia gravis, osteoarthritis, osteoporosis, pancreatitis, psoriasis, viral, bacterial, fungal, parasitic, protozoal or helminthic infections, cancers, autoimmune thyroiditis, cretinism, Plummer's disease or thyroid carcinoma. Embodiments also provide methods for using the purified EXMES and/or their encoding polynucleotides for facilitating the drug discovery process, including determining of efficacy, dosage, toxicity and pharmacology, and for investigating the pathogenesis of diseases and medical conditions.

Seq Sequence 1270 AA;

Query Match 91.8%; Score 5353.7; DB 8; Length 1270;
Best Local Similarity 91.5%; Pred. No. 9.7e-144;
Matches 1006; Conservative 0; Mismatches 0; Indels 93; Gaps 1;

Qy 1 MAWKTLPIYLLLLSVFVIOQVSSQELSCKGRCFESFERGECDCDAQCKYDKCCPDYE 60
Db 1 MAWKTLPIYLLLLSVFVIOQVSSQELSCKGRCFESFERGECDCDAQCKYDKCCPDYE 60

Qy 61 SFCAEVNHTSPSPSKKAPPSPGASQTIKSTTKRSPKPPNKKTKVIESEETEEHSVS 120
Db 61 SFCAE----- 65

Qy 121 ENQSSSSSSSSSSSTIWKSSKNSAANRELQKLKVKDNKNRTRKKPTPKPPVVDE 180
Db 66 -----VKDNKNRTRKKPTPKPPVVDE 87

Qy 181 AGSLGNDGPKVTPPTDSTTQHNVKSTSPKITTAKPINPRPSLPNPSDTSKETSLTVNKE 240
Db 88 AGSLGNDGPKVTPPTDSTTQHNVKSTSPKITTAKPINPRPSLPNPSDTSKETSLTVNKE 147

QY 241 TTVERKETTNTKQSTQDKEKTTSAKETQSIKTSKADLAPTSKVLAKPTPKAETTTKG 300
 Db 148 TTVERKETTNTKQSTQDKEKTTSAKETQSIKTSKADLAPTSKVLAKPTPKAETTTKG 207
 QY 301 PALTTPKEPTTPKEPASTTKEPTTTSKAPTTPKEPAPTTPKSAPTTPKSPAPTTP 360
 Db 208 PALTTPKEPTTPKEPASTTKEPTTTSKAPTTPKEPAPTTPKSAPTTPKSPAPTTP 267
 QY 361 KSPAPTTPKEPAPTTPKSPAPTTPKSAPTTPKSAPTTPKSAPTTPKSAPTTPKSPAPTTP 420
 Db 268 KSPAPTTPKEPAPTTPKSPAPTTPKSAPTTPKSAPTTPKSAPTTPKSAPTTPKSPAPTTP 327
 QY 421 TTPKEPAPTTPKEPAPTTPKSAPTTPKSAPTTPKSAPTTPKSAPTTPKSAPTTPKSPAPTTP 480
 Db 328 TTPKEPAPTTPKEPAPTTPKSAPTTPKSAPTTPKSAPTTPKSAPTTPKSAPTTPKSPAPTTP 387
 QY 481 PAPTTPKSAPTTPKSAPTTPKSAPTTPKSAPTTPKSAPTTPKSAPTTPKSAPTTPKSPAPTTP 540
 Db 388 PAPTTPKSAPTTPKSAPTTPKSAPTTPKSAPTTPKSAPTTPKSAPTTPKSAPTTPKSPAPTTP 447
 QY 541 KSPAPTTPKEPAPTTPKSAPTTPKSAPTTPKSAPTTPKSAPTTPKSAPTTPKSAPTTPKSPAPTTP 600
 Db 448 KSPAPTTPKEPAPTTPKSAPTTPKSAPTTPKSAPTTPKSAPTTPKSAPTTPKSAPTTPKSPAPTTP 507
 QY 601 PTPPELAPTTPEEPTTPPEEPTTPPEEPTTPPEEPTTPPEEPTTPPEEPTTPPEEPTTP 660
 Db 508 PTPPELAPTTPEEPTTPPEEPTTPPEEPTTPPEEPTTPPEEPTTPPEEPTTPPEEPTTP 567
 QY 661 KETAPTTPKGTAPTTPKGTAPTTPKGTAPTTPKGTAPTTPKGTAPTTPKGTAPTTPKGTAPTTP 720
 Db 568 KETAPTTPKGTAPTTPKGTAPTTPKGTAPTTPKGTAPTTPKGTAPTTPKGTAPTTPKGTAPTTP 627
 QY 721 PKAPAPTTPKAPAPTTPKAPAPTTPKAPAPTTPKAPAPTTPKAPAPTTPKAPAPTTPKAPAPTTP 780
 Db 628 PKAPAPTTPKAPAPTTPKAPAPTTPKAPAPTTPKAPAPTTPKAPAPTTPKAPAPTTPKAPAPTTP 687
 QY 781 TPKEAPTTPKEAPTTPKEAPTTPKEAPTTPKEAPTTPKEAPTTPKEAPTTPKEAPTTPKEAPTTP 840
 Db 688 TPKEAPTTPKEAPTTPKEAPTTPKEAPTTPKEAPTTPKEAPTTPKEAPTTPKEAPTTPKEAPTTP 747
 QY 841 ABPTPEALENSKPEGVETTKPATKPEMTTAKDTERDLSTTETTTTAAKMTKET 900
 Db 748 ABPTPEALENSKPEGVETTKPATKPEMTTAKDTERDLSTTETTTTAAKMTKET 807
 QY 901 ATTTEKTTESKITATTQVSTTTQDTPPFKITLKITLAKPVTTTKKTIITTEIMNKP 960
 Db 808 ATTTEKTTESKITATTQVSTTTQDTPPFKITLKITLAKPVTTTKKTIITTEIMNKP 867
 QY 961 BETAKPKDRATNSKATTPKPKOKPKAPKPKPTSTKPKTTPKAVRPKPTTPPKMTSTNPE 1020
 Db 868 BETAKPKDRATNSKATTPKPKOKPKAPKPKPTSTKPKTTPKAVRPKPTTPPKMTSTNPE 927
 QY 1021 LNPTSRIAEAMLQTTTRNQTPNSKLVENPKSEADAGAEETHMLLRHVFMPVETPD 1080
 Db 928 LNPTSRIAEAMLQTTTRNQTPNSKLVENPKSEADAGAEETHMLLRHVFMPVETPD 987
 QY 1081 MDYLPVRVNOGIINPMLS 1099
 Db 988 MDYLPVRVNOGIINPMLS 1006

RESULT 10
 ADK67912

ID ADK67912 standard; protein; 1311 AA.

XX ADK67912;

XX 06-MAY-2004 (first entry)

XX Human extracellular messenger (EXMES) polypeptide.

XX Human; extracellular messenger; EXMES; respiratory-gen.; antiallergic;

KW antiasthmatic; antiinflammatory; antidiabetic; neuroprotective;
 KW muscular-gen.; aniarthritic; osteopathic; hepatotropic; antipsoriatic;
 KW varicide; fungicide; antiparasitic; protozoacide; antihelminthic;
 KW cyostatic; gene therapy.

OS Homo sapiens.

PH Key Location/Qualifiers
 Peptide 1..30
 FT /label= Signal_peptide
 FT /note= "Spans residues 1 to 18, 20, 21, 24, 29 or 30
 according to identification method"

XX WO2004013292-A2.

XX 12-FEB-2004.

XX 30-JUL-2003; 2003WO-US024084.

XX 02-AUG-2002; 2002US-0400810P.

XX 19-SEP-2002; 2002US-0412197P.

XX 04-OCT-2002; 2002US-0416004P.

XX 08-NOV-2002; 2002US-0424862P.

XX (INCY-) INCYTE CORP.

PI Elliott VS, Khare R, Tran UK, Swarnakar A, Marquis JP.

PI Richardson TW, Emerling BM, Lindquist BA, Chawla NK, Ramkumar J;

PI Lee SJ;

XX WPI; 2004-157116/15.

XX N-PSDB; ADK67917.

XX Claim 60; SEQ ID NO 5; 165pp; English.

XX The present sequence is that of novel human extracellular messenger
 (EXMES) Incyte ID NO: 7513018CD1 polypeptide. The protein shows homology
 to human megakaryocyte stimulating factor. The invention provides EXMES
 polynucleotides and polypeptides, as well as expression vectors, host
 cells, antibodies, agonists and antagonists, and methods for diagnosing,
 treating or preventing disorders associated with aberrant expression of
 EXMES, especially autoimmune and inflammatory disorders, cell
 proliferative disorders and endocrine disorders, e.g. adult respiratory
 distress syndrome, allergy, anaemia, erythroblastosis foetalis, Crohn's
 disease, diabetes mellitus, myasthenia gravis, osteoarthritis,
 osteoporosis, pancreatitis, psoriasis, viral, bacterial, fungal,
 parasitic, protozoal or helminthic infections, cancers, autoimmune
 thyroiditis, cretinism, Plummer's disease or thyroid carcinoma.

XX Embodiments also provide methods for using the purified EXMES and/or
 their encoding polynucleotides for facilitating the drug discovery
 process, including determining of efficacy, dosage, toxicity and
 pharmacology, and for investigating the pathogenesis of diseases and
 medical conditions.

XX Sequence 1311 AA;

Query Match 91.5%; Score 5339.6; DB 8; Length 1311;

Best Local Similarity 88.2%; Pred. No. 2.5e-143;

Matches 100%; Conservative 0; Mismatches 0; Indels 134; Gaps 2;

QY 1 MAWKTLPIYLLLSLVFVIQVSSQ----- 25

Db 1 MAWKTLPIYLLLSLVFVIQVSSQDLSSCAGCGEGYSRDATCNDYNCQHYMECCPDF 60

QY 26 -----ELSCKRCPESEFERGECDDAQQKYPKCCPDYESFCAEVNPTSPSSKAP 79

Db 61 KRVTAEILSKRCPESEFERGECDDAQQKYPKCCPDYESFCAE----- 106

QY 80 PPSGASQTIKSTTKRSPKPNKKTKKVISEEBITEHSVSENQESSSSSSSSSTIIV 139
 Db 107 ----- 106
 QY 140 KIKSSKNSAANRELQKLVKXNDKNRKKKPTKPPVWDEAGSLGNDGPKVTPDTST 199
 Db 107 -----VKDNKNRKKKPTKPPVWDEAGSLGNDGPKVTPDTST 147
 QY 200 TOHNKVSTSPKITTAKPINRPSLPNSDTSKETSLTVNKETTVETKETTNNKQTSIDG 259
 Db 148 TOHNKVSTSPKITTAKPINRPSLPNSDTSKETSLTVNKETTVETKETTNNKQTSIDG 207
 QY 260 KEKTSASKEQSIKTSADLAPSKVLAKPTPKAETTTKGPALTTKPEPTPTPKBPAS 319
 Db 208 KEKTSASKEQSIKTSADLAPSKVLAKPTPKAETTTKGPALTTKPEPTPTPKBPAS 267
 QY 320 TTPKEPTTTIKSAPTTKPEAPTTTKSAPTTKPEAPTTTKPEAPTTTKPEAPTTTKPE 379
 Db 268 TTPKEPTTTIKSAPTTKPEAPTTTKSAPTTKPEAPTTTKPEAPTTTKPEAPTTTKPE 327
 QY 380 APTTTKSAPTTKPEAPTTKPKAPTTKPEAPTTKPEPTTTKPEAPTTKPEAPTTPK 439
 Db 328 APTTTKSAPTTKPEAPTTKPKAPTTKPEAPTTKPEPTTTKPEAPTTKPEAPTTPK 387
 QY 440 EPAPTAPKKPAPTTTKPEAPTTKPEAPTTTKPEAPTTTKPEAPTTTKSAPTTKPEAP 499
 Db 398 EPAPTAPKKPAPTTTKPEAPTTKPEAPTTTKPEAPTTTKPEAPTTTKSAPTTKPEAP 447
 QY 500 TTKSAPTTKPEAPTTTKPEAPTTKPEAPTTKPKAPTTKPEAPTTKPEAPTTTKPK 559
 Db 448 TTKSAPTTKPEAPTTTKPEAPTTTKPEAPTTKPKAPTTKPEAPTTTKPEAPTTTKPK 507
 QY 560 APTAPKEAPTTKETAPTTKKLAPTTPEKLAFTTPEKAPTTPEELAPTTPEEPTTT 619
 Db 508 APTAPKEAPTTKETAPTTKKLAPTTPEKLAFTTPEKAPTTPEELAPTTPEEPTTT 567
 QY 620 PEEAPTTPKAAAPNTKPEAPTTKPEAPTTKPEAPTTKPEAPTTKPEAPTTKGTAPTTKEP 679
 Db 568 PEEAPTTPKAAAPNTKPEAPTTKPEAPTTKPEAPTTKPEAPTTKPEAPTTKGTAPTTKEP 627
 QY 680 APTTKKAPKAPKELAPTTTKPEPTSTSDKAPTTPKGTAPTTKPEAPTTKPEAPTTPKG 739
 Db 628 APTTKKAPKAPKELAPTTTKPEPTSTSDKAPTTPKGTAPTTKPEAPTTKPEAPTTPKG 687
 QY 740 TAPTTLKPEAPTTPKKAPKAPKELAPTTTKGPTSTSDKAPTTPKETAFTTKPEAPTTPK 799
 Db 688 TAPTTLKPEAPTTPKKAPKAPKELAPTTTKGPTSTSDKAPTTPKETAFTTKPEAPTTPK 747
 QY 800 KPAPTTPEPTTSEVSTPTTKBPTTIHKSPDESTBELSAEPTPKALENSPKBPGVPT 859
 Db 748 KPAPTTPEPTTSEVSTPTTKBPTTIHKSPDESTBELSAEPTPKALENSPKBPGVPT 807
 QY 860 TKTPAATKPEMTTAKDKTTERDLRTTPETTTAAAPKWTKETATTTTEKTTESKITATTQV 919
 Db 808 TKTPAAKPEMTTAKDKTTERDLRTTPETTTAAAPKWTKETATTTTEKTTESKITATTQV 867
 QY 920 TSTTTQDTTFFKITLLKTTLLAPKVTTTKKLTITTEIMNKPEETAKPKDRATNSKATPK 979
 Db 868 TSTTTQDTTFFKITLLKTTLLAPKVTTTKKLTITTEIMNKPEETAKPKDRATNSKATPK 927
 QY 980 PQKPTKAPKKTPTKPKTMRVRKKTTPTRKWTSTWPELNPTSRIAEAMLQTTTRPN 1039
 Db 928 PQKPTKAPKKTPTKPKTMRVRKKTTPTRKWTSTWPELNPTSRIAEAMLQTTTRPN 987
 QY 1040 QTPNSKLVEVNPKSDEAGGAGETPHMLLRPHVFMPEVTPDMDYLPVRVNOGIIINPMLS 1099
 Db 988 QTPNSKLVEVNPKSDEAGGAGETPHMLLRPHVFMPEVTPDMDYLPVRVNOGIIINPMLS 1047

RESULT 11
 ABUS3252
 ID ABUS3252 standard; protein; 546 AA.
 XX

ABUS3252;
 14-APR-2003 (first entry)
 Human testes-derived DKFZphtes3_4019 homologue #1.
 Human; gene therapy; vaccine; disease treatment; detection.
 Homo sapiens.
 WO200112659-A2.
 22-FEB-2001.
 18-AUG-2000; 2000WO-IB001496.
 18-AUG-1999; 99US-0149499P.
 28-SEP-1999; 99US-0156503P.
 (GEHU-) GERMAN HUMAN GENOME PROJECT.
 Wiemann S;
 WPI; 2001-327840/34.
 Nucleic acids having the sequences of clones isolated from libraries of different human tissues, useful in recombinant DNA methodologies.
 Example III; Page 892; 1095pp; English.
 This invention describes novel polynucleotides and polypeptides isolated from human cDNA libraries which can be used for gene therapy or in vaccines. The polynucleotides of the invention and antibodies encoded by them may be used in the prevention, diagnosis and treatment of diseases associated with inappropriate polypeptide expression. The products of the invention may also be used to identify modulators of expression and activity and to down regulate expression and activity. The antibodies of the invention may also be used as diagnostic agents for detecting the presence of polypeptides in samples. This sequence represents a homologue of a polypeptide described in the disclosure of the invention

Sequence 546 AA;
 Query Match 50.0%; Score 2920; DB 4; Length 546;
 Best Local Similarity 100.0%; Pred. No. 3.5e-75;
 Matches 546; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 447 KKPAFTTPKGPAPTTTKPEAPTTTKPEPTTTKPEAPTTTKSAPTTTKPEAPTTTKSAPT 506
 Db 1 KKPAFTTPKGPAPTTTKPEAPTTTKPEPTTTKPEAPTTTKSAPTTTKPEAPTTTKSAPT 60
 QY 507 TPKEPSPTTTKPEAPTTKPEAPTTPKKAPTTKPEAPTTKPEAPTTTKKAPKAPKE 566
 Db 61 TPKEPSPTTTKPEAPTTKPEAPTTPKKAPTTKPEAPTTKPEAPTTTKKAPKAPKE 120
 QY 567 PAPTTPKETAFTTKPKLTPTTPEKLAFTTPEKAPTTPEELAPTTPEEPTTTPEEAPPT 626
 Db 121 PAPTTPKETAFTTKPKLTPTTPEKLAFTTPEKAPTTPEELAPTTPEEPTTTPEEAPPT 180
 QY 627 TPKAAAPNTKPEAPTTKPEAPTTKPEAPTTPKETAFTTKETAFTTKETAFTTKETAFTTK 686
 Db 181 TPKAAAPNTKPEAPTTKPEAPTTKPEAPTTPKETAFTTKETAFTTKETAFTTKETAFTTK 240
 QY 687 PAPKELAPTTTKETSSTSDKAPTTPKGTAPTTKPEAPTTKPEAPTTKGTAPTTK 746
 Db 241 PAPKELAPTTTKETSSTSDKAPTTPKGTAPTTKPEAPTTKPEAPTTKGTAPTTK 300
 QY 747 EPAPTTPKKAPKAPKELAPTTTKGPTSTSDKAPTTPKETAFTTKPEAPTTPKKAPPTP 806
 Db 301 EPAPTTPKKAPKAPKELAPTTTKGPTSTSDKAPTTPKETAFTTKPEAPTTPKKAPPTP 360
 QY 807 ETPPPTTSEVSTPTTKBPTTIHKSPDESTBELSAEPTPKALENSPKBPGVPTTKTTPAT 866


```

Db 361 ETPTPTTSVSTPTTTTKEPTTIHKSPDESTPELSAETPKALENSPKPEPGVPTTKTPAAT 420
QY 867 KBEMTTAKDKTERDLRTTPTTTAAAPKMTKETATTTTEKTTESKITATTTQVSTTTQD 926
Db 421 KBEMTTAKDKTERDLRTTPTTTAAAPKMTKETATTTTEKTTESKITATTTQVSTTTQD 480
QY 927 TTPFKITLTKTTLAPKVTITTKITTTTTEIMNKPEETAKPKDRATNSKATTPKPKPTKA 986
Db 481 TTPFKITLTKTTLAPKVTITTKITTTTTEIMNKPEETAKPKDRATNSKATTPKPKPTKA 540
QY 987 PKKPTS 992
Db 541 PKKPTS 546

RESULT 12
ABUS3253
ID ABUS3253 standard; protein; 551 AA.
XX AC ABUS3253;
XX DT 14-APR-2003 (first entry)
XX DE Human testes-derived DKFZphtes3_4c19 homologue #2.
XX KW Human; gene therapy; vaccine; disease treatment; detection.
XX OS Homo sapiens.
XX PN WO200112659-A2.
XX PD 22-FEB-2001.
XX PF 18-AUG-2000; 2000WO-IB001496.
XX PR 18-AUG-1999; 99US-0149499P.
XX PR 28-SEP-1999; 99US-0156503P.
XX PA (GEHU-) GERMAN HUMAN GENOME PROJECT.
XX PI Wiemann S;
XX DR WPI; 2001-327840/34.
XX PT Nucleic acids having the sequences of clones isolated from libraries of
XX different human tissues, useful in recombinant DNA methodologies.
XX PS Example III; Page 892-893; 1095pp; English.
XX CC This invention describes novel polynucleotides and polypeptides isolated
XX from human cDNA libraries which can be used for gene therapy or in
XX vaccines. The polynucleotides of the invention and antibodies encoded by
XX them may be used in the prevention, diagnosis and treatment of diseases
XX associated with inappropriate polypeptide expression. The products of the
XX invention may also be used to identify modulators of expression and
XX activity and to down regulate expression and activity. The antibodies of
XX the invention may also be used as diagnostic agents for detecting the
XX presence of polypeptides in samples. This sequence represents a homologue
XX of a polypeptide described in the disclosure of the invention
XX SQ Sequence 551 AA;

Query Match 49.5%; Score 2888; DB 4; Length 551;
Best Local Similarity 100.0%; Pred. No. 2.9e-74;
Matches 544; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 556 TKKPAPAPKEPAPTTPKETAPTPKLTPTTPEKLAPTTPPEKAPAPTTPELAPTTPEEP 615
Db 1 TKKPAPAPKEPAPTTPKETAPTPKLTPTTPEKLAPTTPPEKAPAPTTPELAPTTPEEP 60

QY 616 TPTTPEAPAPTTPKAAAPNTPKSPAPTTPKPEAPTTPKPEAPTTPKETAPTTPKGTAPTT 675
Db 61 TPTTPEAPAPTTPKAAAPNTPKSPAPTTPKPEAPTTPKPEAPTTPKETAPTTPKGTAPTT 120

```

```

QY 676 LKBPAPTPPKKAPKELAPTTTKEPTSTTSKDPAPTPPKGTAPTTPKBPAPTTPKBPAPT 735
Db 121 LKBPAPTPPKKAPKELAPTTTKEPTSTTSKDPAPTPPKGTAPTTPKBPAPTTPKBPAPT 180
QY 736 TPKGTAPTTLKBPAPTPPKKAPKELAPTTTKEPTSTTSKDPAPTTPKBPAPTTPKBPAPT 795
Db 181 TPKGTAPTTLKBPAPTPPKKAPKELAPTTTKEPTSTTSKDPAPTTPKBPAPTTPKBPAPT 240
QY 796 TTPKKAPAPTTPEPTTSEVSTPTTKEPTTIHKSPDESTPELSAETPKALENSPKPEP 855
Db 241 TTPKKAPAPTTPEPTTSEVSTPTTKEPTTIHKSPDESTPELSAETPKALENSPKPEP 300
QY 856 GVPTTKTPAATKPEMTTAKDKTERDLRTTPTTTAAAPKMTKETATTTTEKTTESKITAT 915
Db 301 GVPTTKTPAATKPEMTTAKDKTERDLRTTPTTTAAAPKMTKETATTTTEKTTESKITAT 360
QY 916 TTQVSTTTTQDTPFKITLTKTTLAPKVTITTKITTTTTEIMNKPEETAKPKDRATNSKA 975
Db 361 TTQVSTTTTQDTPFKITLTKTTLAPKVTITTKITTTTTEIMNKPEETAKPKDRATNSKA 420
QY 976 TTPKPKPTKAPKPTSTTKPKTMPRVRKPTTTPRKMVSTMPBLNPTSRIAEAMLOTT 1035
Db 421 TTPKPKPTKAPKPTSTTKPKTMPRVRKPTTTPRKMVSTMPBLNPTSRIAEAMLOTT 480
QY 1036 TRPNQTPNSKLVEVNPKSEDAAGAGETPHMLLRPHVFMPEVTPDMVYLRVFNQGIIN 1095
Db 481 TRPNQTPNSKLVEVNPKSEDAAGAGETPHMLLRPHVFMPEVTPDMVYLRVFNQGIIN 540
QY 1096 PMLS 1099
Db 541 PMLS 544

RESULT 13
AAB29778
ID AAB29778 standard; protein; 902 AA.
XX AC AAB29778;
XX DT 28-FEB-2001 (first entry)
XX DE Human MSF-derived tribonectin.
XX KW Human tribonectin; MSF; megakaryocyte stimulating factor;
XX alternative splicing; joint boundary lubricant; O-linked oligosaccharide;
XX osteoarthritis; tribosupplementation; tissue adhesion inhibition;
XX friction coefficient reduction; gene therapy; antiarthritic; osteopathic.
XX OS Homo sapiens.
XX PN WO200064930-A2.
XX PD 02-NOV-2000.
XX PF 24-APR-2000; 2000WO-US010953.
XX PR 23-APR-1999; 99US-00298970.
XX PA (RHOD-) RHODE ISLAND HOSPITAL LIFESPAN PARTNER.
XX PI Jay GD;
XX DR WPI; 2001-024673/03.
XX PT Novel tribonectin polypeptide useful as lubricant for treating
XX osteoarthritis, comprises O-linked lubricating moiety.
XX PS Disclosure; Fig 1; 47pp; English.
XX CC The invention relates to a human tribonectin which is a product of
XX alternative splicing of the human MSF (megakaryocyte stimulating factor)
XX gene. The tribonectin has at least one O-linked oligosaccharide

```

CC lubricating moiety and has a polypeptide sequence comprising 1-76 repeats
 CC of a motif having at least 50% identity to the sequence KEPAPTT
 CC (AAE29774). The invention also relates to a nucleic acid encoding a human
 CC MSF-derived tribonection; a bio-compatible composition comprising a human
 CC tribonection for inhibiting tissue adhesion formation; and a method of
 CC diagnosing osteoarthritis or a predisposition to osteoarthritis by
 CC measuring the amount of MSF or its fragment in a biological sample of a
 CC mammal, wherein an increased amount of MSF compared to a control
 CC indicates the presence of or predisposition to developing osteoarthritis.
 CC The tribonection and DNA encoding it are useful in the treatment of
 CC osteoarthritis, where they may be used for lubricating mammalian joints,
 CC such as articulating joints of humans, dogs or horses. The tribonection,
 CC when formulated as a membrane, foam, gel or fibre, is useful for
 CC inhibiting adhesion between two surfaces such as the injured tissues of a
 CC mammal where the injury is caused by a surgical insertion or trauma, or
 CC an artificial device e.g., an orthopaedic implant. In particular, one of
 CC the surfaces is pericardial tissue. DNA encoding a tribonection may be
 CC used in gene therapy. The present sequence represents a substantial
 CC portion of a human MSF-derived tribonection
 CC
 CC
 CC SQ Sequence 902 AA;

Query Match 49.4%; Score 2880.1; DB 4; Length 902;
 Best Local Similarity 71.4%; Pred. No. 9.3e-74;
 Matches 675; Conservative 27; Mismatches 74; Indels 169; Gaps 66;

| | | | |
|----|-----|---|-----|
| QY | 159 | VKDNNKQRTKKKPTKPPVVDAGSLDNGDFKVTTPDTSTTQHNKSVTSKITTAKPIN | 218 |
| Db | 1 | VKDNNKQRTKKKPTKPPVVDAGSLDNGDFKVTTPDTSTTQHNKSVTSKITTAKPIN | 60 |
| QY | 219 | PPPSLPNSDTSKETSLSLVNKKETTTNNKQTSDDGKEKTSKETSOSIEKTSK | 278 |
| Db | 61 | PPPSLPNSDTSKETSLSLVNKKETTTNNKQTSDDGKEKTSKETSOSIEKTSK | 120 |
| QY | 279 | DIAPTSKVIAKPTPKAETTTKGPALTTKPEPTTPKPEASTTPKEPTPTTIKSAPTTPK | 338 |
| Db | 121 | DIAPTSKVIAKPTPKAETTTKGPALTTKPEPTTPKPEASTTPKEPTPTTIKSAPTTPK | 172 |
| QY | 339 | EPAPTTTKSAPTTKPEAPTTTKPEAPTTKGPAPTTTKPEAPTTTKSAPTTKPEAPTT | 398 |
| Db | 173 | EPAPTTTKSAPTTKPEAPTTTKPEAPTTTKGPAPTTTKPEAPTTTKSAPTTKPEAPTT | 228 |
| QY | 399 | PKKPAETTKPEAPTTKPEPTTPKPEAPTTKPEAPTTKGPAPTTKPEAPTTKGPAP | 458 |
| Db | 229 | KPEAPTTTKPEAPTTKGPAPTTTKPEAPTTKPEAPTTKGPAPTTKPEAPTTKGPAP | 281 |
| QY | 459 | PTTPKPEAPTTTKPEPTTPKPEAPTTTKSAPTTTKPEAPTTTKSAPTTKPEPTTKKE | 518 |
| Db | 282 | PTT-KEPAP-TTKPEAPTT-KEPAPTTKGPAP-TTKPEAPTTKPEAPTT-KEPAP-TTKE | 335 |
| QY | 519 | PAPTTPKPEAPTTKPEAPTTKPEAPTTKPEAPTTKGPAPTTKGPAPTAPKPEAPTTKETA | 578 |
| Db | 336 | PAPTTP-KEPAPTT-KEPAPTT-KEPAPTT-KEPAPTT-TTKEAPTT-KEPAPTT-KEPAPTT | 388 |
| QY | 579 | TPKLLPTTPEKLAPTTPEKAPTTPEELAPTTPEEPTTTPPEAPTTTPKAAAPNTPK | 638 |
| Db | 389 | T-KEPAPTTKEP-APTKE-PAPTT-KEPAPTT-KEPAPTT-KEPAPTT-KEPAPTT-KE | 440 |
| QY | 639 | PAPTTPKPEAPTTKPEAPTTKPEAPTTKGPAPTTPKGPAPTTPKGPAP--KELAPTT | 596 |
| Db | 441 | PAPTTP-KEPAPTT-KEPAPTT-KEPAPTT-KEPAPTT-KEPAPTT-KEPAPTTKPEAP-T | 493 |
| QY | 697 | TKPEPTSTSDKAPPTPKGAPTTPKPEAPTTTPKPEAPTTKGPAPTTPKPEAPTTTKPKP | 756 |
| Db | 494 | TKGPAPTTKE-PAPTT-KEPAPTT-KEPAPTT-KEPAPTT-KEPAPTT-KEPAPTT-KEP | 546 |
| QY | 757 | AP--KELAPTTTKGPTSTSDKAPPTPKETAAPTTPKPEAPTTKGPAPTTPPTPTTS | 814 |
| Db | 547 | APTTPKPEAP-TTKPEAPTTKE-PAPTT-KEPAPTT-KEPAPTT-KEPAPTTKES-PAPTTK | 600 |
| QY | 815 | EVSTPTTTKEPTTHKSPDSESTELSAPETPKALENSPKPGVPTTKTPAATKPEMTT | 874 |
| Db | 601 | EPA--PTTKGPAPTTKGPAPTT-----KEPAP-----TTKEP-APTKEPAPTT----- | 640 |

| | | | |
|----|------|---|------|
| QY | 875 | KOKTTERDLRTTPTTTAAPKMTKETAATTEKTSKITATTQTQVSTTTTQDTTPPKITT | 934 |
| Db | 641 | -----TKEPAPTT-----KEPAPTTKEPAPTTKEPAP----- | 667 |
| QY | 935 | LKTTTLAPKVTTTKTITTTIMNNKEETAKEPKDRATNSKATTPKPKTKAPKFTSTK | 994 |
| Db | 668 | -----TTKEPAPTTKEP-APT-TK | 684 |
| QY | 995 | KPKTMRVRKPKTTPTRKMTSTMPELNPTSRIAEAMLQTTTRPNQTPNSKLVENVNPKGE | 1054 |
| Db | 685 | EP-----APTTPRKMSTMPELNPTSRIAEAML-TTTRPNQTPNSKLVENVNPKGE | 735 |
| QY | 1055 | DAGGAGETPHMLLRPHVMPVETTPDMVLPVNOGIIINPMLS | 1099 |
| Db | 736 | DAGGAGETPHMLLRPHVMPVETTPDMVLPVNOGIIINPMLS | 780 |

RESULT 14

ABUS3254
 ID ABUS3254 standard; protein; 513 AA.

| | | |
|----|---|---------------|
| XX | AC | ABUS3254; |
| XX | AC | |
| DT | 14-APR-2003 | (first entry) |
| DE | Human testes-derived DKFZphtes3_4019 | homologue #3. |
| XX | Human; gene therapy; vaccine; disease treatment; detection. | |
| OS | Homo sapiens. | |
| XX | WO2000112659-A2. | |
| PD | 22-FEB-2001. | |
| XX | 18-AUG-2000; 2000WO-IB001496. | |
| PR | 18-AUG-1999; 99US-0149499P. | |
| PR | 28-SEP-1999; 99US-0156503P. | |
| XX | (GEHU-) GERMAN HUMAN GENOME PROJECT. | |
| XX | Wiemann S; | |
| XX | WPI; 2001-327840/34. | |

Nucleic acids having the sequences of clones isolated from libraries of different human tissues, useful in recombinant DNA methodologies.

Example III; Page 893; 1095pp; English.

This invention describes novel polynucleotides and polypeptides isolated from human cDNA libraries which can be used for gene therapy or in vaccines. The polynucleotides of the invention and antibodies encoded by them may be used in the prevention, diagnosis and treatment of diseases associated with inappropriate polypeptide expression. The products of the invention may also be used to identify modulators of expression and activity and to down regulate expression and activity. The antibodies of the invention may also be used as diagnostic agents for detecting the presence of polypeptides in samples. This sequence represents a homologue of a polypeptide described in the disclosure of the invention

Sequence 513 AA;

Query Match 47.2%; Score 2757; DB 4; Length 513;

Best Local Similarity 100.0%; Pred. No. 1.4e-70;

Matches 513; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

| | | | |
|----|-----|---|-----|
| QY | 166 | RTKKKPTKPPVVDAGSLDNGDFKVTTPDTSTTQHNKSVTSKITTAKPINRPSLPP | 225 |
| Db | 1 | RTKKKPTKPPVVDAGSLDNGDFKVTTPDTSTTQHNKSVTSKITTAKPINRPSLPP | 60 |

QY 226 NSDTSKETSLSLVNKEVETVETKETTNNKQSTDGKETTSAKETQSIKTSARDLAPTSK 285
 DB 61 NSDTSKETSLSLVNKEVETVETKETTNNKQSTDGKETTSAKETQSIKTSARDLAPTSK 120
 QY 286 VLAKPTPKASTTTKGPALTTPKSPTPTTPKEPASTTPKPTPTTIKSAPTTPKEPAPTTT 345
 DB 121 VLAKPTPKASTTTKGPALTTPKSPTPTTPKEPASTTPKPTPTTIKSAPTTPKEPAPTTT 180
 QY 346 KSAPTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKSAPTTPKEPAPTTTPKPPAPT 405
 DB 181 KSAPTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKSAPTTPKEPAPTTTPKPPAPT 240
 QY 406 TPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKSAPTTPKEPAPTTTPKPPAPT 465
 DB 241 TPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKSAPTTPKEPAPTTTPKPPAPT 300
 QY 466 APTTTPKEPAPTTTPKEPAPTTTPKSAPTTPKEPAPTTTPKSAPTTPKEPAPTTTPK 525
 DB 301 APTTTPKEPAPTTTPKEPAPTTTPKSAPTTPKEPAPTTTPKSAPTTPKEPAPTTTPK 360
 QY 526 EPAPTTPKPPAPTTPKEPAPTTTPKSAPTTPKEPAPTTTPKSAPTTPKEPAPTTTPK 585
 DB 361 EPAPTTPKPPAPTTPKEPAPTTTPKSAPTTPKEPAPTTTPKSAPTTPKEPAPTTTPK 420
 QY 586 TTPEKLAFTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKSAPTTPKEPAPTTTPK 645
 DB 421 TTPEKLAFTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKSAPTTPKEPAPTTTPK 480
 QY 646 EPAPTTPKEPAPTTTPKEPAPTTTPKSAPTTPKEPAPTTTPKSAPTTPKEPAPTTTPK 678
 DB 481 EPAPTTPKEPAPTTTPKEPAPTTTPKSAPTTPKEPAPTTTPKSAPTTPKEPAPTTTPK 513

RESULT 15

AA080041
 ID AAR80041 standard; protein; 452 AA.

XX AAR80041;
 AC AAR80041;

XX 25-MAR-2003 (revised)
 DT 10-APR-1996 (first entry)

XX Human megakaryocytopoietin protein.

XX Human; megakaryocytopoietin; wheat germ agglutinin; heparin;
 KW megakaryocyte; aplastic anaemia; thrombocytopenia; hematocytopenia;
 KW multipotential stem cell.

XX Homo sapiens.

XX Key Location/Qualifiers
 FH Misc-difference 393..396
 FT /note= "unspecified amino acids"

FT Misc-difference 444..446
 FT /note= "unspecified amino acids"

XX W09523861-A1.

XX 08-SEP-1995.

XX 06-MAR-1995; 95WO-CN000015.

XX 04-MAR-1994; 94CN-00112066.

XX (SHAN-) SHANGHAI BEITE BIOTECHNOLOGY CO LTD.

XX Gu X, Han Z, Shen Q;

XX WPI; 1995-320576/41.

XX N-PSDB; AAT04545.

XX New haematopoietic cell growth factor - used for treating
 PT thrombocytopenia and hematocytopenia.

XX Example; Page 23; 36pp; Chinese.

XX This sequence represents the human megakaryocytopoietin (MPO) protein.
 CC This sequence was purified using a carrier which can couple wheat germ
 CC agglutinin and heparin to separate MPO. Fragments of this sequence (see
 CC AAR80039 and AAR80040) were used to produce the amplification primers
 CC shown in AAT04544 and AAT04545. The fragments amplified by these primers
 CC can then be used as probes to screen human cDNA libraries for MPO cDNA.
 CC The MPO cDNA can then be inserted into a plasmid which is used to
 CC transform cells to produce MPO. The MPO sequence is capable of promoting
 CC colony formation of megakaryocytes, enlarging the size of megakaryocytes
 CC and stimulating the proliferation of multipotential stem cells. The
 CC factor may be used for treating thrombocytopenia and hematocytopenia. The
 CC purification method can be used to isolate MPO from human urine or serum
 CC of patients with aplastic anaemia, and from animal blood or urine by
 CC radiation exposing the animals to induce aplastic anaemia. (Updated on 25
 CC -MAR-2003 to correct PA field.)

XX Sequence 452 AA;

Query Match 33.5%; Score 1954.5; DB 2; Length 452;
 Best Local Similarity 76.1%; Pred. NO. 6.8e-48;

Matches 378; Conservative 3; Mismatches 21; Indels 95; Gaps 3;

QY 1 MAWKTLPIVLLLLSVFVIQVSSQ----- 25

DB 1 MAWKTLPIVLLLLSVFVIQVSSQDLSSCAGRCGSGYSRDATCNDYNCQHYMECCPDF 60

QY 26 -----ELCKGRCPESFERGECDCDAQCKYDKCCPDYESFCAEVHNFTSPSSKKAP 79

DB 61 KRVTAEELCKGRCPESFERGECDCDAQCKYDKCCPDYESFCAEVHNFTSPSSKKAP 120

QY 80 PPSGASQTIKSTTKGSPKPPNKKTKKVIKESBEITEHSVSENQSSSSSSSSSTIW 139

DB 121 PPSGASQTIKSTTKGSPKPPNKKTKKVIKESBEITE----- 156

QY 140 KIKSSKNSAANRELQKLVKONKQNRKTKKTPKPPVVVDGAGSLDNGDFKVTPTDST 199

DB 157 -----VKDNKNRTKKKTPKPPVVVDGAGSLDNGDFKVTPTDST 197

QY 200 TQHNKVSTSPKITTAKPINPRSLPNSDTSKETSLSLVNKEVETVETKETTNNKQSTDG 259

DB 198 TQHNKVSTSPKITTAKPINPRSLPNSDTSKETSLSLVNKEVETVETKETTNNKQSTDG 257

QY 260 KEKTTSAKETQSIKTSARDLAPTSKVLAKPTPKAETTTKGPALTTTPKEPTTPKPEAS 319

DB 258 KEKTTSAKETQSIKTSARDLAPTSKVLAKPTPKAETTTKGPALTTTPKEPTTPKPEAS 317

QY 320 TTPKEPTPTTIKSAPTTPKEPAPTTTKSAPTTTPKEPAPTTTPKSAPTTPKEPAPTTTPK 379

DB 318 TTPKEPTPTTIKSAPTTPKEPAPTTTKSAPTTTPKEPAPTTTPKSAPTTPKEPAPTTTPK 377

QY 380 APTTTPKSAPTTPKEPAPTTTPKPPAPTTPKSAPTTPKEPAPTTTPKSAPTTPKEPAPTTTPK 439

DB 378 APTTTPKSHPPPLPRSCXXXCTQP-----TPKEPHPLPRSLHPTTPKEPAPTTPK 426

QY 440 EPAPTAPKPKKAPPTTPE 456

DB 427 EPAPTAPKPKKAPLPPLPE 443

Search completed: October 13, 2004, 11:36:40
 Job time : 96.6045 secs

cl1 (no alignment)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: October 13, 2004, 11:23:49 ; Search time 19.1039 Seconds
(without alignments)
3615.116 Million cell updates/sec

Title: SEQ1-A
Perfect score: 5835
Sequence: 1 NAWKTLPRYLLLLSVFVIQ.....DMDYLPVPGIINPMLS 1099

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.1

Searched: 478139 seqs, 66318000 residues

Total number of hits satisfying chosen parameters: 478139

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued Patents AA:*

- 1: /cgn2_6/prodata/1/aaa/5A-COMB.pep:*
- 2: /cgn2_6/prodata/1/aaa/5B-COMB.pep:*
- 3: /cgn2_6/prodata/1/aaa/5A-COMB.pep:*
- 4: /cgn2_6/prodata/1/aaa/5B-COMB.pep:*
- 5: /cgn2_6/prodata/1/aaa/PCTUS-COMB.pep:*
- 6: /cgn2_6/prodata/1/aaa/backfiles1.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | ID | Description |
|------------|--------|-------------|--------|----|--------------------|
| 1 | 5835 | 100.0 | 1363 | 4 | US-07-757-022B-52 |
| 2 | 5820.9 | 99.8 | 1140 | 4 | US-07-757-022B-104 |
| 3 | 5820.9 | 99.8 | 1404 | 4 | US-07-757-022B-2 |
| 4 | 5820.9 | 99.8 | 1404 | 4 | US-07-757-022B-62 |
| 5 | 5820.9 | 99.8 | 1404 | 4 | US-09-298-970A-1 |
| 6 | 5797.9 | 99.4 | 1404 | 4 | US-10-164-595-78 |
| 7 | 5617.7 | 96.3 | 1320 | 4 | US-07-757-022B-46 |
| 8 | 5617.7 | 96.3 | 1320 | 4 | US-07-757-022B-60 |
| 9 | 5608.7 | 96.1 | 1320 | 4 | US-10-164-595-58 |
| 10 | 5603.6 | 96.0 | 1361 | 4 | US-07-757-022B-40 |
| 11 | 5561 | 95.3 | 1049 | 4 | US-07-757-022B-58 |
| 12 | 5561 | 95.3 | 1313 | 4 | US-07-757-022B-142 |
| 13 | 5546.9 | 95.1 | 1354 | 4 | US-07-757-022B-48 |
| 14 | 5377.9 | 93.2 | 1314 | 4 | US-07-757-022B-50 |
| 15 | 5353.7 | 91.8 | 1038 | 4 | US-07-757-022B-74 |
| 16 | 5353.7 | 91.5 | 1270 | 4 | US-07-757-022B-44 |
| 17 | 5339.6 | 91.5 | 1311 | 4 | US-07-757-022B-84 |
| 18 | 5331.9 | 89.7 | 1022 | 4 | US-07-757-022B-14 |
| 19 | 5011 | 85.9 | 941 | 4 | US-09-538-092-1258 |
| 20 | 1354.4 | 23.2 | 5179 | 4 | US-07-757-022B-96 |
| 21 | 1331.8 | 19.4 | 891 | 4 | US-08-714-741-32 |
| 22 | 923.9 | 15.8 | 220 | 4 | US-07-757-022B-96 |
| 23 | 864.6 | 14.8 | 207 | 4 | US-07-757-022B-116 |
| 24 | 864.6 | 14.8 | 207 | 4 | US-07-757-022B-136 |
| 25 | 862.2 | 14.8 | 422 | 4 | US-07-757-022B-94 |
| 26 | 861.9 | 14.8 | 209 | 4 | US-07-757-022B-68 |
| 27 | 856.9 | 14.7 | 208 | 4 | US-07-757-022B-132 |

| | | | | | |
|----|-------|------|------|---|--------------------|
| 28 | 855.7 | 14.7 | 3256 | 4 | US-09-919-172-98 |
| 29 | 855.7 | 14.7 | 3256 | 4 | US-09-976-594-22 |
| 30 | 855.7 | 14.7 | 3256 | 4 | US-09-919-039-21 |
| 31 | 848.1 | 14.5 | 463 | 4 | US-07-757-022B-54 |
| 32 | 845.7 | 14.5 | 2972 | 3 | US-09-579-181-2 |
| 33 | 845.7 | 14.5 | 3118 | 3 | US-09-579-181-1 |
| 34 | 835.9 | 14.3 | 204 | 4 | US-07-757-022B-92 |
| 35 | 823.1 | 14.1 | 296 | 4 | US-07-757-022B-70 |
| 36 | 811.5 | 13.9 | 231 | 4 | US-07-757-022B-30 |
| 37 | 792.1 | 13.6 | 2142 | 4 | US-09-538-092-1142 |
| 38 | 785.9 | 13.5 | 1837 | 3 | US-08-928-361B-5 |
| 39 | 785.9 | 13.5 | 1837 | 4 | US-09-588-995A-5 |
| 40 | 776.9 | 13.3 | 192 | 4 | US-07-757-022B-90 |
| 41 | 744.4 | 12.8 | 1721 | 3 | US-08-700-651-5 |
| 42 | 743.4 | 12.7 | 1721 | 3 | US-08-928-361B-6 |
| 43 | 743.4 | 12.7 | 1721 | 4 | US-09-588-995A-6 |
| 44 | 734.1 | 12.6 | 423 | 4 | US-07-757-022B-66 |
| 45 | 727.9 | 12.5 | 3969 | 3 | US-08-061-376-5 |

ALIGNMENTS

RESULT 1
US-07-757-022B-52
; Sequence 52, Application US/07757022B
; Patent No. 6433142
; GENERAL INFORMATION:
; APPLICANT: Gesner, Thomas G.
; APPLICANT: Clark, Stephen C.
; APPLICANT: Turner, Katherine
; APPLICANT: Hewick, Rodney M.
; TITLE OF INVENTION: Megakaryocyte Stimulating Factors
; NUMBER OF SEQUENCES: 143
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genetics Institute, Inc.
; STREET: 87 Cambridgepark Drive
; CITY: Cambridge
; STATE: Massachusetts
; COUNTRY: U.S.A.
; ZIP: 02140

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/07757,022B
FILING DATE: 19910910
CLASSIFICATION: 530
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/643,502
FILING DATE: 18-JAN-1991
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/546,114
FILING DATE: 29-JUN-1990
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/457,196
FILING DATE: 29-DEC-1989
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/390,901
FILING DATE: 08-AUG-1989
ATTORNEY/AGENT INFORMATION:
NAME: Ceerr, Luann
REGISTRATION NUMBER: 31,822
REFERENCE/DOCKET NUMBER: GI 5190
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617)876-1170
TELEFAX: (617)876-5851
INFORMATION FOR SEQ ID NO: 52:
SEQUENCE CHARACTERISTICS:
LENGTH: 1363 amino acids
TYPE: AMINO ACID

```

; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-07-757-022B-52

```

Query Match 100.0%; Score 5835; DB 4; Length 1363;
Best Local Similarity 100.0%; Pred. No. 4.5e-170;
Matches 1099; Conservative 0; Mismatches 0; Indels 0;

| | | | |
|----|-----|--|------|
| QY | 1 | MAWKTLPIYLLLLSVFVIOQVSSQELSCKGRCFESFERGBECDDAQCKYKDKCCPYE | 60 |
| Db | 1 | MAWKTLPIYLLLLSVFVIOQVSSQELSCKGRCFESFERGBECDDAQCKYKDKCCPYE | 60 |
| QY | 61 | SFCAEVHNPTGPPSSKAPPPSGASQIKSTTKRSPKPNKKTKKVIIESEIITEHSVS | 120 |
| Db | 61 | SFCAEVHNPTGPPSSKAPPPSGASQIKSTTKRSPKPNKKTKKVIIESEIITEHSVS | 120 |
| QY | 121 | ENQESSSSSSSSSSSIWKIKSQGSAANREIQKLVKDNKNRTKKPTPKPPVUDE | 180 |
| Db | 121 | ENQESSSSSSSSSSSIWKIKSSKNSAANREIQKLVKDNKNRTKKPTPKPPVUDE | 180 |
| QY | 181 | AGSGLDNGDFKVTTPDSTTTOHNKVSTSPKITTAKPINRPSLPNPSDKETSILTVAKE | 240 |
| Db | 181 | AGSGLDNGDFKVTTPDSTTTOHNKVSTSPKITTAKPINRPSLPNPSDKETSILTVAKE | 240 |
| QY | 241 | TTVETKETITNNKOTSDGKEKTTSAKETOSI EKTSAKOLAPTSVKLAKPTPKABTTKG | 300 |
| Db | 241 | TTVETKETITNNKOTSDGKEKTTSAKETOSI EKTSAKOLAPTSVKLAKPTPKABTTKG | 300 |
| QY | 301 | PALITPKBPTTTTPEKPASTPKBPTPTTIKSAPTTTKBPAPTTTKSAPTTKBPAPTTT | 360 |
| Db | 301 | PALITPKBPTTTTPEKPASTPKBPTPTTIKSAPTTTKBPAPTTTKSAPTTKBPAPTTT | 360 |
| QY | 361 | KEPAPTTPKBAPTTTKEPAPTTTKSAPTTKBPAPTTTKBPAPTTPKBAPTTPKBPTP | 420 |
| Db | 361 | KEPAPTTPKBAPTTTKEPAPTTTKSAPTTKBPAPTTTKBPAPTTPKBAPTTPKBPTP | 420 |
| QY | 421 | TTPEKAPTTKEPAPTTPKBPAPTPAPKPKAPTTTKEBPAPTTTKEBPAPTTKEBPTTKE | 480 |
| Db | 421 | TTPEKAPTTKEPAPTTPKBPAPTPAPKPKAPTTTKEBPAPTTTKEBPAPTTKEBPTTKE | 480 |
| QY | 481 | PAPTTTKSAPTTTKEPAPTTTKSAPTTPKBPSPTTTKEPAPTTPKBPAPTTPKBPAPTT | 540 |
| Db | 481 | PAPTTTKSAPTTTKEPAPTTTKSAPTTPKBPSPTTTKEPAPTTPKBPAPTTPKBPAPTT | 540 |
| QY | 541 | KEPAPTTPKBPAPTTTKKBPAPTAKBPAPTTPKETAPTTPKLTPTTPEKLPAPTTPKPA | 600 |
| Db | 541 | KEPAPTTPKBPAPTTTKKBPAPTAKBPAPTTPKETAPTTPKLTPTTPEKLPAPTTPKPA | 600 |
| QY | 601 | PTTPEELAPTTPEBPTTTTPEBPAPTTPKAAPNTPKBPAPTTPKBPAPTTPKBPAPTT | 660 |
| Db | 601 | PTTPEELAPTTPEBPTTTTPEBPAPTTPKAAPNTPKBPAPTTPKBPAPTTPKBPAPTT | 660 |
| QY | 661 | KETAPTTPKGTAPTTLKEPAPTTPKBPAPTTKBPAPTTTSDKBPAPTTPKGTAPTT | 720 |
| Db | 661 | KETAPTTPKGTAPTTLKEPAPTTPKBPAPTTKBPAPTTTSDKBPAPTTPKGTAPTT | 720 |
| QY | 721 | PKBPAPTTPKBPAPTTPKGTAPTTLKEPAPTTPKBPAPTTTSGFTSTTSDKAPTT | 780 |
| Db | 721 | PKBPAPTTPKBPAPTTPKGTAPTTLKEPAPTTPKBPAPTTTSGFTSTTSDKAPTT | 780 |
| QY | 781 | TPKETAPTTKEPAPTTPKBPAPTTPEBPTTSEVSTPTTTKEPTTIHKSPDSESTELS | 840 |
| Db | 781 | TPKETAPTTKEPAPTTPKBPAPTTPEBPTTSEVSTPTTTKEPTTIHKSPDSESTELS | 840 |
| QY | 841 | AEPTPKALENSPKBPGVPTTKTTPAAKTPBMTTAKDKTERDLRTPPTTTTAAKPMTKET | 900 |
| Db | 841 | AEPTPKALENSPKBPGVPTTKTTPAAKTPBMTTAKDKTERDLRTPPTTTTAAKPMTKET | 900 |
| QY | 901 | ATTTEKTTESKITATTTOVTSITTTODTTPFKLITLKTITLAPKVTTTKKITTITTEIMNKP | 960 |
| Db | 901 | ATTTEKTTESKITATTTOVTSITTTODTTPFKLITLKTITLAPKVTTTKKITTITTEIMNKP | 960 |
| QY | 961 | SEETAKPDRAITNGKATTPKQKETKAPKPTSTKPKPTMPRVKPKTTPTPRKMTSTWPE | 1020 |

```

Db      961  EETAKPKDRATNSKATTPKPKQKTKAPKPKTSTKPKKTKTPRVEKPKTTPTPRKWTSTMPD 1020
QY      1021  LNPSTSIARLAMLQTTTRPQNTNSKLVEVNPKSEDAGGAGETPHMLLRPHVFMPEVTPD 1080
Db      1021  LNPSTSIARLAMLQTTTRPQNTNSKLVEVNPKSEDAGGAGETPHMLLRPHVFMPEVTPD 1080
QY      1081  MDYLPRVPNQGIINPMLS 1099
Db      1081  MDYLPRVPNQGIINPMLS 1099

RESULT 2
US-07-757-022B-104
; Sequence 104, Application US/07757022B
; Patent No. 6433142
; GENERAL INFORMATION:
; APPLICANT: Gesner, Thomas G.
; APPLICANT: Clark, Stephen C.
; APPLICANT: Turner, Katherine
; APPLICANT: Hewick, Rodney M.
; TITLE OF INVENTION: Megakaryocyte Stimulating Factors
; NUMBER OF SEQUENCES: 143
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genetics Institute, Inc.
; STREET: 87 Cambridgepark Drive
; CITY: Cambridge
; STATE: Massachusetts
; COUNTRY: U.S.A.
; ZIP: 02140
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/757,022B
; FILING DATE: 19910910
; CLASSIFICATION: 530
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/643,502
; FILING DATE: 18-JAN-1991
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/546,114
; FILING DATE: 29-JUN-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/457,196
; FILING DATE: 29-DEC-1989
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/390,901
; FILING DATE: 08-AUG-1989
; ATTORNEY/AGENT INFORMATION:
; NAME: Ceert, Luann
; REGISTRATION NUMBER: 31,822
; REFERENCE/DOCKET NUMBER: GI 5190
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617) 876-1170
; TELEFAX: (617) 876-5851
; INFORMATION FOR SEQ ID NO: 104:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1140 amino acids
; TYPE: AMINO ACID
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-07-757-022B-104

Query Match          99.8%; Score 5820.9; DB 4; Length 1140;
Best Local Similarity 96.4%; Pred. No. 9.7e-170;
Matches 1099; Conservative 0; Mismatches 0; Indels 41; Gaps 1;

QY      1 MAWKTLPIYLLLSLVFVIQVSSQ----- 25
Db      1 MAWKTLPIYLLLSLVFVIQVSSQDILSSCAGRGEGYSRDATCNCDYNCOHMECCPDF 60

```

QY 26 -----ELSCRGCFESFERGECDDAQCKYDKCCPDYSEFCAEVHNFTSPSSKKAP 79
DB 61 KRVCTAELSCRGCFESFERGECDDAQCKYDKCCPDYSEFCAEVHNFTSPSSKKAP 120
QY 80 PPSGASQTIKSTTKSPKPNKKTKKVIIESEEITEHVSSENGESSSSSSSSSTIWI 139
DB 121 PPSGASQTIKSTTKSPKPNKKTKKVIIESEEITEHVSSENGESSSSSSSSSTIWI 180
QY 140 KIKSSKNSAANRELOKLVKNDKNKRTKKKPTPKPPVVDAGSLDNGDFKVTTPDTST 199
DB 181 KIKSSKNSAANRELOKLVKNDKNKRTKKKPTPKPPVVDAGSLDNGDFKVTTPDTST 240
QY 200 TQHNKVSTSPKITTAKPINRPSLPNSDTSKETSILVKNKETTIVETKETTINKQITDGTG 259
DB 241 TQHNKVSTSPKITTAKPINRPSLPNSDTSKETSILVKNKETTIVETKETTINKQITDGTG 300
QY 260 KEKTTSAKETQSIKTSKADLAPTSKVLAKEPTTKGPAETTTKGPALTTKPEPTTTPKEPAS 319
DB 301 KEKTTSAKETQSIKTSKADLAPTSKVLAKEPTTKGPAETTTKGPALTTKPEPTTTPKEPAS 360
QY 320 TTPKEPTTTTIKASPTTKPEAPATTGAPTTKPEAPATTGAPTTKPEAPATTGAPTTKEP 379
DB 361 TTPKEPTTTTIKASPTTKPEAPATTGAPTTKPEAPATTGAPTTKPEAPATTGAPTTKEP 420
QY 380 APTTTKSAPTTTPKEAPATTGAPTTKPEAPATTGAPTTKPEAPATTGAPTTKEP 439
DB 421 APTTTKSAPTTTPKEAPATTGAPTTKPEAPATTGAPTTKPEAPATTGAPTTKEP 480
QY 440 EPAPAPKAPATTTPKEAPATTGAPTTKPEAPATTGAPTTKPEAPATTGAPTTKEP 499
DB 481 EPAPAPKAPATTTPKEAPATTGAPTTKPEAPATTGAPTTKPEAPATTGAPTTKEP 540
QY 500 TTKSAPTTTPKEAPATTGAPTTKPEAPATTGAPTTKPEAPATTGAPTTKEP 559
DB 541 TTKSAPTTTPKEAPATTGAPTTKPEAPATTGAPTTKPEAPATTGAPTTKEP 600
QY 560 APTAPKEAPATTTPKEAPATTGAPTTKPEAPATTGAPTTKPEAPATTGAPTTKEP 619
DB 601 APTAPKEAPATTTPKEAPATTGAPTTKPEAPATTGAPTTKPEAPATTGAPTTKEP 660
QY 620 PEPAPATTTPKAAPTTTPKEAPATTGAPTTKPEAPATTGAPTTKPEAPATTGAPTTKEP 679
DB 661 PEPAPATTTPKAAPTTTPKEAPATTGAPTTKPEAPATTGAPTTKPEAPATTGAPTTKEP 720
QY 680 APTTPKAPKAPLAPTTTPKEAPATTGAPTTKPEAPATTGAPTTKPEAPATTGAPTTKEP 739
DB 721 APTTPKAPKAPLAPTTTPKEAPATTGAPTTKPEAPATTGAPTTKPEAPATTGAPTTKEP 780
QY 740 TAPTTLKAPAPTTTPKAPKAPLAPTTTPKEAPATTGAPTTKPEAPATTGAPTTKPEAPATT 799
DB 781 TAPTTLKAPAPTTTPKAPKAPLAPTTTPKEAPATTGAPTTKPEAPATTGAPTTKPEAPATT 840
QY 800 KPAPTTTPPTPTTSEVSTPTTTKEPTTIHKSPDESTPELSAEPPTKALENSPKPEPVPT 859
DB 841 KPAPTTTPPTPTTSEVSTPTTTKEPTTIHKSPDESTPELSAEPPTKALENSPKPEPVPT 900
QY 860 TKTPAAKTEMTTAKDKTTERDLRTTPTTTPAAKMTKETATTTKTESKITATTQV 919
DB 901 TKTPAAKTEMTTAKDKTTERDLRTTPTTTPAAKMTKETATTTKTESKITATTQV 960
QY 920 TSTTTTODTTPPKITTLKTTLAPKVTITTKTITTEIMNKPETAKPKDRATNSKATTPK 979
DB 961 TSTTTTODTTPPKITTLKTTLAPKVTITTKTITTEIMNKPETAKPKDRATNSKATTPK 1020
QY 980 POKPTKAPKPTSTKPKTTPMVRKPTTTPRKMSTSTWPELNPTSRIAEAMLOTTTPN 1039
DB 1021 POKPTKAPKPTSTKPKTTPMVRKPTTTPRKMSTSTWPELNPTSRIAEAMLOTTTPN 1080
QY 1040 QTPNSKLVNPKSEADAGAEGETPHMLRPHVFPVETPDMDYLPVFNQGIINPMLS 1099
DB 1081 QTPNSKLVNPKSEADAGAEGETPHMLRPHVFPVETPDMDYLPVFNQGIINPMLS 1140

RESULT 3
US-07-757-022B-2
; Sequence 2, Application US/07757022B
; Patent No. 6433142
; GENERAL INFORMATION:
; APPLICANT: Gesner, Thomas G.
; APPLICANT: Clark, Stephen C.
; APPLICANT: Turner, Katherine
; APPLICANT: Hewick, Rodney M.
; TITLE OF INVENTION: Megakaryocyte Stimulating Factors
; NUMBER OF SEQUENCES: 143
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genetics Institute, Inc.
; STREET: 87 CambridgePark Drive
; CITY: Cambridge
; STATE: Massachusetts
; COUNTRY: U.S.A.
; ZIP: 02140
; COMPUTER READABLE FORM: disk
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/757,022B
; FILING DATE: 19910910
; CLASSIFICATION: 530
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/643,502
; FILING DATE: 18-JAN-1991
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/546,114
; FILING DATE: 29-JUN-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/457,196
; FILING DATE: 29-DEC-1989
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/390,901
; FILING DATE: 08-AUG-1989
; ATTORNEY/AGENT INFORMATION:
; NAME: Cseri, Luann
; REGISTRATION NUMBER: 31,822
; REFERENCE/DOCKET NUMBER: G1 5190
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617)876-1170
; TELEFAX: (617)876-5851
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1404 amino acids
; TYPE: AMINO ACID
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-07-757-022B-2
Query Match 99.8%; Score 5820.9; DB 4; Length 1404;
Best Local Similarity 96.4%; Pred. No. 1.3e-169;
Matches 1099; Conservative 0; Mismatches 0; Indels 41; Gaps 1;
QY 1 MANKTLPIYLLLSLVFIQVSSQ----- 25
DB 1 MANKTLPIYLLLSLVFIQVSSQ----- 60
QY 26 -----ELSCRGCFESFERGECDDAQCKYDKCCPDYSEFCAEVHNFTSPSSKKAP 79
DB 61 KRVCTAELSCRGCFESFERGECDDAQCKYDKCCPDYSEFCAEVHNFTSPSSKKAP 120
QY 80 PPSGASQTIKSTTKSPKPNKKTKKVIIESEEITEHVSSENGESSSSSSSSSTIWI 139
DB 121 PPSGASQTIKSTTKSPKPNKKTKKVIIESEEITEHVSSENGESSSSSSSSSTIWI 180
QY 140 KIKSSKNSAANRELOKLVKNDKNKRTKKKPTPKPPVVDAGSLDNGDFKVTTPDTST 199

Db 181 KIKSSKNSAANRELQKLVKNDKNKRTKKKPTKPPVVDGAGSLDNGDFKVTTPDTST 240
 QY 200 TQHNKVSTSPKITTAKPINRPSLDPNSDTSKETSLSLVNKTETVETKETTNTKQTSIDG 259
 Db 241 TQHNKVSTSPKITTAKPINRPSLDPNSDTSKETSLSLVNKTETVETKETTNTKQTSIDG 300
 QY 260 KEKTTSAKETQSIKTSKADLAPTSKVLAKPTPKAETTTKGPALTTKPEPTTTPKBPAS 319
 Db 301 KEKTTSAKETQSIKTSKADLAPTSKVLAKPTPKAETTTKGPALTTKPEPTTTPKBPAS 360
 QY 320 TTPKESPTPTTIKSAPTTPKBPAPTTTKSAPTTTKBPAPTTTKBPAPTTTKBP 379
 Db 361 TTPKESPTPTTIKSAPTTPKBPAPTTTKSAPTTTKBPAPTTTKBPAPTTTKBP 420
 QY 380 APTTTKSAPTTPKBPAPTTTKBPAPTTTKBPAPTTTKBPAPTTTKBPAPTTTKBP 439
 Db 421 APTTTKSAPTTPKBPAPTTTKBPAPTTTKBPAPTTTKBPAPTTTKBPAPTTTKBP 480
 QY 440 EPAPTAPKBPAPTTTKBPAPTTTKBPAPTTTKBPAPTTTKBPAPTTTKBPAPTT 499
 Db 481 EPAPTAPKBPAPTTTKBPAPTTTKBPAPTTTKBPAPTTTKBPAPTTTKBPAPTT 540
 QY 500 TTKSAPTTTKBPAPTTTKBPAPTTTKBPAPTTTKBPAPTTTKBPAPTTTKBPAPTT 559
 Db 541 TTKSAPTTTKBPAPTTTKBPAPTTTKBPAPTTTKBPAPTTTKBPAPTTTKBPAPTT 600
 QY 560 APTAPKBPAPTTTKBPAPTTTKBPAPTTTKBPAPTTTKBPAPTTTKBPAPTTTKBP 619
 Db 601 APTAPKBPAPTTTKBPAPTTTKBPAPTTTKBPAPTTTKBPAPTTTKBPAPTTTKBP 660
 QY 620 PEEAPPTPKAAANPTPKBPAPTTTKBPAPTTTKBPAPTTTKBPAPTTTKBPAPTTTKBP 679
 Db 661 PEEAPPTPKAAANPTPKBPAPTTTKBPAPTTTKBPAPTTTKBPAPTTTKBPAPTTTKBP 720
 QY 680 APTPKKAPKELAPTTTKBPAPTTTKBPAPTTTKBPAPTTTKBPAPTTTKBPAPTTTKBP 739
 Db 721 APTPKKAPKELAPTTTKBPAPTTTKBPAPTTTKBPAPTTTKBPAPTTTKBPAPTTTKBP 780
 QY 740 TAPTTKBPAPTTTKBPAPTTTKBPAPTTTKBPAPTTTKBPAPTTTKBPAPTTTKBPAPTT 799
 Db 781 TAPTTKBPAPTTTKBPAPTTTKBPAPTTTKBPAPTTTKBPAPTTTKBPAPTTTKBPAPTT 840
 QY 800 KPAPTPTPTPTSEVSTPTTKBPAPTTTKBPAPTTTKBPAPTTTKBPAPTTTKBPAPTTTKBP 859
 Db 841 KPAPTPTPTPTSEVSTPTTKBPAPTTTKBPAPTTTKBPAPTTTKBPAPTTTKBPAPTTTKBP 900
 QY 860 TKTPAATKPEMTTAKDXTTERDLRTPTPTTAAAPKMTKETATTTKTSKITATTTQV 919
 Db 901 TKTPAATKPEMTTAKDXTTERDLRTPTPTTAAAPKMTKETATTTKTSKITATTTQV 960
 QY 920 TSTTTQDTPPKITTLKTTILAPKVTTTKITTTTEIMNKPEETAKPKDRAHNSKATTPK 979
 Db 961 TSTTTQDTPPKITTLKTTILAPKVTTTKITTTTEIMNKPEETAKPKDRAHNSKATTPK 1020
 QY 980 POKTAPKPKTSTKPKIMPRVKPKTTPTPRKMTSTMPKELNPTSRIAEAMQTTTRPN 1039
 Db 1021 POKTAPKPKTSTKPKIMPRVKPKTTPTPRKMTSTMPKELNPTSRIAEAMQTTTRPN 1080
 QY 1040 QTPNSKLVEVNPKGEDAGGAGETPHMLLRPHVPMPEVTPDMDYLPVNPQGIINPMLS 1099
 Db 1081 QTPNSKLVEVNPKGEDAGGAGETPHMLLRPHVPMPEVTPDMDYLPVNPQGIINPMLS 1140

RESULT 4

US-07-757-022B-62
 ; Sequence 62, Application US/07757022B
 ; Patent No. 6433142
 ; GENERAL INFORMATION:
 ; APPLICANT: Gesner, Thomas G.
 ; APPLICANT: Clark, Stephen C.
 ; APPLICANT: Turner, Katherine
 ; APPLICANT: Hewick, Rodney M.
 ; TITLE OF INVENTION: Megakaryocyte Stimulating Factors

NUMBER OF SEQUENCES: 143
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: Genetics Institute, Inc.
 STREET: 87 CambridgePark Drive
 CITY: Cambridge
 STATE: Massachusetts
 COUNTRY: U.S.A.
 ZIP: 02140
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: Patent In Release #1.0, Version #1.25
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/07/757,022B
 FILING DATE: 19910910
 CLASSIFICATION: 530
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 07/643,502
 FILING DATE: 18-JAN-1991
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 07/546,114
 FILING DATE: 29-JUN-1990
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 07/457,196
 FILING DATE: 29-DEC-1989
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 07/390,901
 FILING DATE: 08-AUG-1989
 ATTORNEY/AGENT INFORMATION:
 NAME: Cserr, Luann
 REGISTRATION NUMBER: 31,822
 REFERENCE/DOCKET NUMBER: GI 5190
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (617) 876-1170
 TELEFAX: (617) 876-5851
 INFORMATION FOR SEQ ID NO: 62:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 1404 amino acids
 TYPE: AMINO ACID
 TOPOLOGY: linear
 MOLECULE TYPE: protein
 US-07-757-022B-62

Query Match 99.8%; Score 5820.9; DB 4; Length 1404;
 Best Local Similarity 96.4%; Pred. No. 1.3e-169;
 Matches 1099; Conservative 0; Mismatches 0; Indels 41; Gaps 1;

QY 1 MAWKTLPIYLLLLLVFVIQVSSQ----- 25
 Db 1 MAWKTLPIYLLLLLVFVIQVSSQDLSSCAGRCGEGYRDATCNDYNCQHYMECCDF 60
 QY 26 -----ELSCKRCGFESFERGECDCDAQCKYDKCCPDYEFCAEVHNPTSPSSSKAP 79
 Db 61 KRVCTAELSKRCGFESFERGECDCDAQCKYDKCCPDYEFCAEVHNPTSPSSSKAP 120
 QY 80 PPSGASQIKSTTKSPKPPNKKTKKVIIESEETEEHSVSENESSSSSSSSSSSTI 139
 Db 121 PPSGASQIKSTTKSPKPPNKKTKKVIIESEETEEHSVSENESSSSSSSSSSSTI 180
 QY 140 KIKSSKNSAANRELQKLVKNDKNKRTKKKPTKPPVVDGAGSLDNGDFKVTTPDTST 199
 Db 181 KIKSSKNSAANRELQKLVKNDKNKRTKKKPTKPPVVDGAGSLDNGDFKVTTPDTST 240
 QY 200 TQHNKVSTSPKITTAKPINRPSLDPNSDTSKETSLSLVNKTETVETKETTNTKQTSIDG 259
 Db 241 TQHNKVSTSPKITTAKPINRPSLDPNSDTSKETSLSLVNKTETVETKETTNTKQTSIDG 300
 QY 260 KEKTTSAKETQSIKTSKADLAPTSKVLAKPTPKAETTTKGPALTTKPEPTTTPKBPAS 319
 Db 301 KEKTTSAKETQSIKTSKADLAPTSKVLAKPTPKAETTTKGPALTTKPEPTTTPKBPAS 360
 QY 320 TTPKESPTPTTIKSAPTTPKBPAPTTTKSAPTTTKBPAPTTTKBPAPTTTKBP 379

QY 1040 QTPNSKLVEVNPKESEDAGAGETPHMLLRPHVFMPEVTPDMDYLPVFNQGIINPMLS 1099
 Db 1081 QTPNSKLVEVNPKESEDAGAGETPHMLLRPHVFMPEVTPDMDYLPVFNQGIINPMLS 1140

RESULT 6
 US-10-164-595-78
 ; Sequence 78, Application US/10164595
 ; Patent No. 6657054
 ; GENERAL INFORMATION:
 ; APPLICANT: OriGene Technologies, Inc
 ; TITLE OF INVENTION: Regulated Angiogenesis Genes and Polypeptides
 ; FILE REFERENCE: IU 103 R1
 ; CURRENT APPLICATION NUMBER: US/10/164,595
 ; CURRENT FILING DATE: 2002-06-10
 ; NUMBER OF SEQ ID NOS: 80
 ; SOFTWARE: Patent in version 3.1
 ; SEQ ID NO 78
 ; LENGTH: 1404
 ; TYPE: PRT
 ; ORGANISM: Homo sapiens
 US-10-164-595-78

Query Match 99.4%; Score 5797.9; DB 4; Length 1404;
 Best Local Similarity 96.1%; Pred. No. 6.3e-169;
 Matches 1096; Conservative 0; Mismatches 3; Indels 41; Gaps 1;

QY 1 MAWKTLPIYLLLLSVFVIQVSSQ----- 25
 Db 1 MAWKTLPIYLLLLSVFVIQVSSQDLSSCAGRCGYSRDATCNCYNOCHNMECCPDF 60

QY 26 -----ELSCRGRCFESFERGECDCDAQCKYDKCCPDYDFCAEVHNPTSPSSKKAP 79
 Db 61 KAVCTAELSCRGRCFESFERGECDCDAQCKYDKCCPDYDFCAEVHNPTSPSSKKAP 120

QY 80 PPSGASQIKSTTKESPPNKKTKKVIIEBETEEHSVSENGESESSESSSSSTIIV 139
 Db 121 PPSGASQIKSTTKRSPPNKKTKKVIIEBETEEHSVSENGESESSESSSSSTIR 180

QY 140 KIKSSKNAANRELQKLVKNDKNKRTKKXPTKPPVWDAGSLDNGDFKVTTPDST 199
 Db 181 KIKSSKNAANRELQKLVKNDKNKRTKKXPTKPPVWDAGSLDNGDFKVTTPDST 240

QY 200 TOHNKVSTSPKITTAKPINRPSLPNDSKTSKSLTVNKETTVETKETTNNKQTSIDG 259
 Db 241 TOHNKVSTSPKITTAKPINRPSLPNDSKTSKSLTVNKETTVETKETTNNKQTSIDG 300

QY 260 KEKTTSAKETOSIEKTSADLAPSKVLAKPTPKAETTKGPAITTPKEPTTPKEPAS 319
 Db 301 KEKTTSAKETOSIEKTSADLAPSKVLAKPTPKAETTKGPAITTPKEPTTPKEPAS 360

QY 320 TTPKEPTTTIKSAPTTPKEPATTTKSAPTTPKEPATTTTPKEPATTTTPKEPAT 379
 Db 361 TTPKEPTTTIKSAPTTPKEPATTTKSAPTTPKEPATTTTPKEPATTTTPKEPAT 420

QY 380 APTTKSAPTTPKEPATTPKKEPATTPKEPATTPKEPATTPKEPATTPKEPATTPK 439
 Db 421 APTTKSAPTTPKEPATTPKKEPATTPKEPATTPKEPATTPKEPATTPKEPATTPK 480

QY 440 EPAPTAPKPAPTTPKEPATTPKEPATTTTKESPTTPKEPATTTTKSAPTTTKEPAT 499
 Db 481 EPAPTAPKPAPTTPKEPATTPKEPATTTTKESPTTPKEPATTTTKSAPTTTKEPAT 540

QY 500 TTKSAPTTPKEPSPTTKKEPATTPKKEPATTPKKEPATTPKKEPATTPKKEPATTKK 559
 Db 541 TTKSAPTTPKEPSPTTKKEPATTPKKEPATTPKKEPATTPKKEPATTPKKEPATTKK 600

QY 560 APTAPKEPATTPKETAPTTPKLTPTTPKEKLAPTTPKEPATTPPELAPTTPEEPTPT 619
 Db 601 APTTPKEPATTPKETAPTTPKLTPTTPKEKLAPTTPKEPATTPPELAPTTPEEPTPT 660

QY 620 PEEPAPTPKAAANTPKBPAPPTTPKEPATTPKETAPTTPKGTAPTTKBPAPTTKBP 679

Db 661 PEEPAPTPKAAANTPKBPAPPTTPKEPATTPKETAPTTPKGTAPTTKBPAPTTKBP 720
 QY 680 APTTPKBPAPKELAPTTTKETPTSTTSKBPAPTTPKGTAPTTKBPAPTTKBPAPTTK 739
 Db 721 APTTPKBPAPKELAPTTTKETPTSTTSKBPAPTTPKGTAPTTKBPAPTTKBPAPTTK 780

QY 740 TAPTTLKEBPAPTTPKBPAPKELAPTTTKGTPTSTTSKBPAPTTKETAPTTPKEPATTPK 799
 Db 781 TAPTTLKEBPAPTTPKBPAPKELAPTTTKGTPTSTTSKBPAPTTKETAPTTPKEPATTPK 840

QY 800 KPAPTTPETPTTSEVSTPTTTKEPTTIHKSPDSTPELSAETPKALENSPKPVGPT 859
 Db 841 KPAPTTPETPTTSEVSTPTTTKEPTTIHKSPDSTPELSAETPKALENSPKPVGPT 900

QY 860 TKTPAATKPEMTITAKDKTTERDLRTTPTTAAAPMKETATTTTEKTESKITATTTQV 919
 Db 901 TKTPAATKPEMTITAKDKTTERDLRTTPTTAAAPMKETATTTTEKTESKITATTTQV 960

QY 920 TSTTTQDTPPEKLTTLTKTTLAPKVTITTKTITTTIMNKPBEAKPKDRATNSKATTPK 979
 Db 961 TSTTTQDTPPEKLTTLTKTTLAPKVTITTKTITTTIMNKPBEAKPKDRATNSKATTPK 1020

QY 980 PQKPTKAPKKPTSTKKPKTMRVRKPTTPTRKMTSTMPELNPTSRIAEAMLOTTTRPN 1039
 Db 1021 PQKPTKAPKKPTSTKKPKTMRVRKPTTPTRKMTSTMPELNPTSRIAEAMLOTTTRPN 1080

QY 1040 QTPNSKLVEVNPKESEDAGAGETPHMLLRPHVFMPEVTPDMDYLPVFNQGIINPMLS 1099
 Db 1081 QTPNSKLVEVNPKESEDAGAGETPHMLLRPHVFMPEVTPDMDYLPVFNQGIINPMLS 1140

RESULT 7
 US-07-757-022B-46
 ; Sequence 46, Application US/07757022B
 ; Patent No. 6433142
 ; GENERAL INFORMATION:
 ; APPLICANT: Gesner, Thomas G.
 ; APPLICANT: Clark, Stephen C.
 ; APPLICANT: Turner, Katherine
 ; APPLICANT: Hewick, Rodney M.
 ; TITLE OF INVENTION: Megakaryocyte Stimulating Factors
 ; NUMBER OF SEQUENCES: 143
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Genetics Institute, Inc.
 ; STREET: 87 CambridgePark Drive
 ; CITY: Cambridge
 ; STATE: Massachusetts
 ; COUNTRY: U.S.A.
 ; ZIP: 02140
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: Patent In Release #1.0, Version #1.25
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/07/757,022B
 ; FILING DATE: 19910910
 ; CLASSIFICATION: 530
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: US 07/643,502
 ; FILING DATE: 18-JAN-1991
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: US 07/546,114
 ; FILING DATE: 29-JUN-1990
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: US 07/457,196
 ; FILING DATE: 29-DEC-1989
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: US 07/390,901
 ; FILING DATE: 08-AUG-1989
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Cserr, Luann
 ; REGISTRATION NUMBER: 31,822

```

; REFERENCE/DOCKET NUMBER: GI 5190
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617)876-1170
; TELEFAX: (617)876-5851
; INFORMATION FOR SEQ ID NO: 46:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1320 amino acids
; TYPE: AMINO ACID
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-07-757-022B-46

Query Match          96.3%; Score 5617.7; DB 4; Length 1320;
Best Local Similarity 96.1%; Pred. No. 1.8e-163;
Matches 1056; Conservative 0; Mismatches 0; Indels 43; Gaps 1;

QY 1 MAWKLPYLLLLLVFIQVSSOELSCGRCFSPERGRECDCAOCKYDKCCPDYE 60
DB 1 MAWKLPYLLLLLVFIQVSSOELSCGRCFSPERGRECDCAOCKYDKCCPDYE 60

QY 61 SFCAEVHNPTSPSSKAPPPSGASQTIKSTTKRSPKPPNKKTKKVIKIESEIIEHSVS 120
DB 61 SFCAEVHNPTSPSSKAPPPSGASQTIKSTTKRSPKPPNKKTKKVIKIESEIIE 115

QY 121 ENQESSSSSSSSSTIWKIKSSKNSAANRELQKLVKUNKKORTKKKTPKPPVDE 180
DB 116 -----VKONKKORTKKKTPKPPVDE 137

QY 181 AGSGLDNGDFKVTTPDSTTQHNKYSTSPKITTAKPINRPSLPNSDTSKETSITVNKE 240
DB 138 AGSGLDNGDFKVTTPDSTTQHNKYSTSPKITTAKPINRPSLPNSDTSKETSITVNKE 197

QY 241 TVVETKETTITNKQISTDCGKETTSAKETQSEKTSAKDLAPTSKVLAKTPKASTTKG 300
DB 198 TVVETKETTITNKQISTDCGKETTSAKETQSEKTSAKDLAPTSKVLAKTPKASTTKG 257

QY 301 PALTTPEKPTTPKPEASTTKEPTTIKSAPTTPKEPATTPKSAPTTPKPEAPTTT 360
DB 258 PALTTPEKPTTPKPEASTTKEPTTIKSAPTTPKPEPATTPKSAPTTPKPEAPTTT 317

QY 361 KEPAPTTPKEPATTTTKEPATTTKSAPTTTKEPATTPKKEPATTPKPEAPTTTKEPTP 420
DB 318 KEPAPTTPKEPATTTTKEPATTTKSAPTTTKEPATTPKKEPATTPKPEAPTTTKEPTP 377

QY 421 TTPKEPATTPKEPATTPKPEAPTTAPKPAPTTPKEPATTPKPEAPTTTKEPSTTPKE 480
DB 378 TTPKEPATTPKPEPATTPKPEAPTTAPKPAPTTPKPEAPTTTKEPSTTPKE 437

QY 481 PAPTITKSAPTTTKEPATTTKSAPTTTKEPATTPKPEPATTPKPEAPTTTKEPATTP 540
DB 438 PAPTITKSAPTTTKEPATTTKSAPTTTKEPATTPKPEPATTPKPEAPTTTKEPATTP 497

QY 541 KEPAPTTPKEPATTTTKEPATTPKPEPATTPKPEPATTPKPEPATTPKPEPATTP 600
DB 498 KEPAPTTPKEPATTTTKEPATTPKPEPATTPKPEPATTPKPEPATTPKPEPATTP 557

QY 601 PTTPEELAPTTPEEPTTPEEAPTTTPEEAPTTTPEEAPTTTPEEAPTTTPEEAPTTT 660
DB 558 PTTPEELAPTTPEEPTTPEEAPTTTPEEAPTTTPEEAPTTTPEEAPTTTPEEAPTTT 617

QY 661 KETAPTTKGTAPTTTKEPATTTKPKPAKELAPTTTKGFTSTSDKPAPTPKGAPATT 720
DB 618 KETAPTTKGTAPTTTKEPATTTKPKPAKELAPTTTKGFTSTSDKPAPTPKGAPATT 677

QY 721 PKEPATTPKPEPATTPKGTAPTTTKEPATTTKPKPAKELAPTTTKGFTSTSDKPAPT 780
DB 678 PKEPATTPKPEPATTPKGTAPTTTKEPATTTKPKPAKELAPTTTKGFTSTSDKPAPT 737

QY 781 TPKETAPTTKPEPATTPKPKAPATTTPETPPPTTSEVSTPTTTKEPTTIHKSPDESTPELS 840
DB 738 TPKETAPTTKPEPATTPKPKAPATTTPETPPPTTSEVSTPTTTKEPTTIHKSPDESTPELS 797

QY 841 AEPTPKALENSPKPEGVPTTKTTPAATKPEMTTAKOKITTERDLRTTPEITTAAPKMTKET 900

```

```

RESULT 8
US-07-757-022B-60
; Sequence 60, Application US/07/57022B
; Patent No. 6433142
; GENERAL INFORMATION:
; APPLICANT: Gesner, Thomas G.
; APPLICANT: Clark, Stephen C.
; APPLICANT: Turner, Katherine
; APPLICANT: Hewick, Rodney M.
; TITLE OF INVENTION: Megakaryocyte Stimulating Factors
; NUMBER OF SEQUENCES: 143
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genetics Institute, Inc.
; STREET: 87 Cambridgepark Drive
; CITY: Cambridge
; STATE: Massachusetts
; COUNTRY: U.S.A.
; ZIP: 02140
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/757,022B
; FILING DATE: 19910910
; CLASSIFICATION: 530
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/643,502
; FILING DATE: 18-JAN-1991
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/546,114
; FILING DATE: 29-JUN-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/457,196
; FILING DATE: 29-DEC-1989
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/390,901
; FILING DATE: 08-AUG-1989
; ATTORNEY/AGENT INFORMATION:
; NAME: Csezz, Luann
; REGISTRATION NUMBER: 31,822
; REFERENCE/DOCKET NUMBER: GI 5190
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617)876-1170
; TELEFAX: (617)876-5851
; INFORMATION FOR SEQ ID NO: 60:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1320 amino acids
; TYPE: AMINO ACID
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-07-757-022B-60

```

```

Query Match      96.3%; Score 5617.7; DB 4; Length 1320;
Best Local Similarity 96.1%; Pred. No. 1.8e-163;
Matches 1056; Conservative 0; Mismatches 0; Indels 43; Gaps 1;

QY 1 MAWKTLPIVLLLLSVFVIQVSSQBELSKGRCFSEFGRGECDDAQCCKYDKCCPDYE 60
DB 1 MAWKTLPIVLLLLSVFVIQVSSQBELSKGRCFSEFGRGECDDAQCCKYDKCCPDYE 60

QY 61 SFCAEVHNTSPSSKAPPPSGASQTIKSTTKRSPKPNKKTKKVIIESEBITEHSVS 120
DB 61 SFCAEVHNTSPSSKAPPPSGASQTIKSTTKRSPKPNKKTKKVIIESEBITE- 115
QY 121 ENQESSSSSSSSSTIWKIKSSKNSAANRELQKKLVKONKKKRTKKKPTPKPPVDE 180
DB 116 -----VKONKKKRTKKKPTPKPPVDE 137

QY 181 AGSGLDNGDFKVTDTSTTOHNVSTSPKITTAKPINPRPSLPNSDTSKETSILTVNKE 240
DB 138 AGSGLDNGDFKVTDTSTTOHNVSTSPKITTAKPINPRPSLPNSDTSKETSILTVNKE 197

QY 241 TTVEKKEITTTNKQSTDKGKTTSAKETQSIEKTSAKDLAPTSKVLAKPTKKASTTTKG 300
DB 198 TTVEKKEITTTNKQSTDKGKTTSAKETQSIEKTSAKDLAPTSKVLAKPTKKASTTTKG 257

QY 301 PALTTPKEPTTTPKEPASTTTPKEPTTTIKSAPTTKPEPAPTTTKSAPTTPKPAPTTT 360
DB 258 PALTTPKEPTTTPKEPASTTTPKEPTTTIKSAPTTKPEPAPTTTKSAPTTPKPAPTTT 317

QY 361 KEPAATTPKEPAPTTTKPEPAPTTTKSAPTTKPEPAPTTTKPAPTTPKPAPTTPKG 420
DB 318 KEPAATTPKEPAPTTTKPEPAPTTTKSAPTTKPEPAPTTTKPAPTTPKPAPTTPKG 377

QY 421 TTPKEPAPTTTKPEPAPTTTKPEPAPTTAPKAPAPTTTKPEPAPTTTKPEPAPTTTKE 480
DB 378 TTPKEPAPTTTKPEPAPTTTKPEPAPTTAPKAPAPTTTKPEPAPTTTKPEPAPTTTKE 437

QY 481 PAPTTPKSGAPTTPKSGAPTTPKSGAPTTPKSGAPTTPKSGAPTTPKSGAPTTPKSGAPT 540
DB 438 PAPTTPKSGAPTTPKSGAPTTPKSGAPTTPKSGAPTTPKSGAPTTPKSGAPTTPKSGAPT 497

QY 541 KEPAATTPKEPAPTTTKPEPAPTTTKPEPAPTTTKPEPAPTTTKPEPAPTTTKPEPAP 600
DB 498 KEPAATTPKEPAPTTTKPEPAPTTTKPEPAPTTTKPEPAPTTTKPEPAPTTTKPEPAP 557

QY 601 PTTPEELAPTTPEEPTTTPKEPAPTTPKAAAPNTPKPEPAPTTTPKEPAPTTTPKEPAP 660
DB 558 PTTPEELAPTTPEEPTTTPKEPAPTTPKAAAPNTPKPEPAPTTTPKEPAPTTTPKEPAP 617

QY 661 KETAPTTKSGAPTTPKPEPAPTTTKKAPKELAPTTTKEPTSTTSDDKAPTTPKGTA 720
DB 618 KETAPTTKSGAPTTPKPEPAPTTTPKPKAPKELAPTTTKEPTSTTSDDKAPTTPKGTA 677

QY 721 PKEPAPTTPKPEPAPTTPKGTAAPTTLKPEPAPTTPKKAPKELAPTTTKEPTSTTSDDKAP 780
DB 678 PKEPAPTTPKPEPAPTTPKGTAAPTTLKPEPAPTTPKKAPKELAPTTTKEPTSTTSDDKAP 737

QY 781 TPKEPAPTTPKPEPAPTTPKKAPAPTTPEPTTSEVSTPTTKEPTTIHKSPDSEBTELS 840
DB 738 TPKEPAPTTPKPEPAPTTPKKAPAPTTPEPTTSEVSTPTTKEPTTIHKSPDSEBTELS 797

QY 841 AEPTPKALENSPKPGVPTTKTAPATPEMTTAKDKTTERDLRTTTPETTTAAAPKMTKET 900
DB 798 AEPTPKALENSPKPGVPTTKTAPATPEMTTAKDKTTERDLRTTTPETTTAAAPKMTKET 857

QY 901 ATTTEKTTESKITAATTQVSTTTQDTPPKITLKTTLTTLAPKVTITTKKTIITTEIMNKP 960
DB 858 ATTTEKTTESKITAATTQVSTTTQDTPPKITLKTTLTTLAPKVTITTKKTIITTEIMNKP 917

QY 961 EETAKPKDRATNSKATTPKPKOKPTKAPKPTSTKPKTMRVRKPKTTPPKMTSTMP 1020
DB 918 EETAKPKDRATNSKATTPKPKOKPTKAPKPTSTKPKTMRVRKPKTTPPKMTSTMP 977

```

```

QY 1021 INPSTRIAEAMLOTTTRPNQTPNSKLVENVNPKSEDAGGAEGETPHMLLRPHVFMPEVTPD 1080
DB 978 INPSTRIAEAMLOTTTRPNQTPNSKLVENVNPKSEDAGGAEGETPHMLLRPHVFMPEVTPD 1037

QY 1081 MDYLPVRVNOGIIINPMLS 1099
DB 1038 MDYLPVRVNOGIIINPMLS 1056

RESULT 9
US-10-164-595-58
; Sequence 58, Application US/10164595
; Patent No. 6657054
; GENERAL INFORMATION:
; APPLICANT: Origene Technologies, Inc
; TITLE OF INVENTION: Regulated Angiogenesis Genes and Polypeptides
; FILE REFERENCE: IU 103 R1
; CURRENT APPLICATION NUMBER: US/10/164,595
; CURRENT FILING DATE: 2002-06-10
; NUMBER OF SEQ ID NOS: 80
; SOFTWARE: Patencin version 3.1
; SEQ ID NO 58
; LENGTH: 1320
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-164-595-58

```

```

Query Match      96.1%; Score 5608.7; DB 4; Length 1320;
Best Local Similarity 95.9%; Pred. No. 3.4e-163;
Matches 1054; Conservative 0; Mismatches 2; Indels 43; Gaps 1;

QY 1 MAWKTLPIVLLLLSVFVIQVSSQBELSKGRCFSEFGRGECDDAQCCKYDKCCPDYE 60
DB 1 MAWKTLPIVLLLLSVFVIQVSSQBELSKGRCFSEFGRGECDDAQCCKYDKCCPDYE 60

QY 61 SFCAEVHNTSPSSKAPPPSGASQTIKSTTKRSPKPNKKTKKVIIESEBITEHSVS 120
DB 61 SFCAEVHNTSPSSKAPPPSGASQTIKSTTKRSPKPNKKTKKVIIESEBITE- 115
QY 121 ENQESSSSSSSSSTIWKIKSSKNSAANRELQKKLVKONKKKRTKKKPTPKPPVDE 180
DB 116 -----VKONKKKRTKKKPTPKPPVDE 137

QY 181 AGSGLDNGDFKVTDTSTTOHNVSTSPKITTAKPINPRPSLPNSDTSKETSILTVNKE 240
DB 138 AGSGLDNGDFKVTDTSTTOHNVSTSPKITTAKPINPRPSLPNSDTSKETSILTVNKE 197

QY 241 TTVEKKEITTTNKQSTDKGKTTSAKETQSIEKTSAKDLAPTSKVLAKPTKKASTTTKG 300
DB 198 TTVEKKEITTTNKQSTDKGKTTSAKETQSIEKTSAKDLAPTSKVLAKPTKKASTTTKG 257

QY 301 PALTTPKEPTTTPKEPASTTTPKEPTTTIKSAPTTKPEPAPTTTKSAPTTPKPAPTTT 360
DB 258 PALTTPKEPTTTPKEPASTTTPKEPTTTIKSAPTTKPEPAPTTTKSAPTTPKPAPTTT 317

QY 361 KEPAATTPKEPAPTTTKPEPAPTTTKSAPTTKPEPAPTTTKPAPTTPKPAPTTPKG 420
DB 318 KEPAATTPKEPAPTTTKPEPAPTTTKSAPTTKPEPAPTTTKPAPTTPKPAPTTPKG 377

QY 421 TTPKEPAPTTTKPEPAPTTTKPEPAPTTAPKAPAPTTTKPEPAPTTTKPEPAPTTTKE 480
DB 378 TTPKEPAPTTTKPEPAPTTTKPEPAPTTAPKAPAPTTTKPEPAPTTTKPEPAPTTTKE 437

QY 481 PAPTTPKSGAPTTPKSGAPTTPKSGAPTTPKSGAPTTPKSGAPTTPKSGAPTTPKSGAPT 540
DB 438 PAPTTPKSGAPTTPKSGAPTTPKSGAPTTPKSGAPTTPKSGAPTTPKSGAPTTPKSGAPT 497

QY 541 KEPAATTPKEPAPTTTKPEPAPTTTKPEPAPTTTKPEPAPTTTKPEPAPTTTKPEPAP 600
DB 498 KEPAATTPKEPAPTTTKPEPAPTTTKPEPAPTTTKPEPAPTTTKPEPAPTTTKPEPAP 557

QY 601 PTTPEELAPTTPEEPTTTPKEPAPTTPKAAAPNTPKPEPAPTTTPKEPAPTTTPKEPAP 660

```

558 PTTPEELAPTTPEETPTTPEBPAPTTTAKAAAPNTKPEPAPTTTKEPAPTTTKEPAPTTT 617
661 KETAPTTTKEGTAATTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEGTAATTT 720
618 KETAPTTTKEGTAATTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEGTAATTT 677
721 KPEPAPTTTKEPAPTTTKEGTAATTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEGTAATTT 780
678 KPEPAPTTTKEPAPTTTKEGTAATTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEGTAATTT 737
781 TPEKAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTT 840
738 TPEKAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTT 797
841 APTTTPKALENSKEGVPPTTKPAATTKEMTTTAKDKTERDLRTTPTTPTTAAAPKMTKET 900
798 APTTTPKALENSKEGVPPTTKPAATTKEMTTTAKDKTERDLRTTPTTPTTAAAPKMTKET 857
901 ATTTEKTTESKITATTQTQVSTTTTQDTPPKTTLTKTTTLPKVTTKKLTITTEIMNKP 960
858 ATTTEKTTESKITATTQTQVSTTTTQDTPPKTTLTKTTTLPKVTTKKLTITTEIMNKP 917
961 ESTAKPKDRAATSKATTPKPKQPTKAPKPKPTSTKPKTMPRVRKPKTTTPTPKMTSTWPE 1020
918 ESTAKPKDRAATSKATTPKPKQPTKAPKPKPTSTKPKTMPRVRKPKTTTPTPKMTSTWPE 977
1021 LNPTSRIAEAMLTQTTTRPNQTPNSKLVEVNPKSEDAGGAEGETPHMLLRPHVFMPEVTPD 1080
978 LNPTSRIAEAMLTQTTTRPNQTPNSKLVEVNPKSEDAGGAEGETPHMLLRPHVFMPEVTPD 1037
1081 MDYLPVRVNOGIIINPMLS 1099
1038 MDYLPVRVNOGIIINPMLS 1056

RESULT 10

US-07-757-022B-40
Sequence 40, Application US/07757022B
Patent No. 6433142
GENERAL INFORMATION:
APPLICANT: Gesner, Thomas G.
APPLICANT: Clark, Stephen C.
APPLICANT: Turner, Katherine
APPLICANT: Hewick, Rodney M.
TITLE OF INVENTION: Megakaryocyte Stimulating Factors
NUMBER OF SEQUENCES: 143
CORRESPONDENCE ADDRESS:
ADDRESSEE: Genetics Institute, Inc.
STREET: 87 CambridgePark Drive
CITY: Cambridge
STATE: Massachusetts
COUNTRY: U.S.A.
ZIP: 02140
COMPUTER READABLE FORM: disk
MEDIUM TYPE: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/07/757,022B
FILING DATE: 19910910
CLASSIFICATION: 530
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/643,502
FILING DATE: 18-JAN-1991
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/546,114
FILING DATE: 29-JUN-1990
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/457,196
FILING DATE: 29-DEC-1989
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/390,901

FILING DATE: 08-AUG-1989
ATTORNEY/AGENT INFORMATION:
NAME: Cserr, Luann
REGISTRATION NUMBER: 31,822
REFERENCE/DOCKET NUMBER: GI 5190
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617)876-1170
TELEFAX: (617)876-5851
INFORMATION FOR SEQ ID NO: 40:
SEQUENCE CHARACTERISTICS:
LENGTH: 1361 amino acids
TYPE: AMINO ACID
TOPOLOGY: linear
MOLECULE TYPE: protein
US-07-757-022B-40
Query Match 96.0%; Score 5603.6; DB 4; Length 1361;
Best Local Similarity 92.6%; Pred. No. 5e-163;
Matches 1056; Conservative 0; Mismatches 0; Indels 84; Gaps 2;
QY 1 MANKTLPIYLLLSLVFIQQVSSQ----- 25
DB 1 MANKTLPIYLLLSLVFIQQVSSQDLSSCAGCGEYSRDATAICNDYNOQHMECCPDF 60
QY 26 -----ELSCCKGRCFESFERGECDDAOCKKYDKCCPDYESFCAEVHNFTSPSSKKAP 79
DB 61 KRVTALSLCKGRCFESFERGECDDAOCKKYDKCCPDYESFCAEVHNFTSPSSKKAP 120
QY 80 PPGASOTIKSTTKRSPKPNKKTKKVISEEITEHSVSENOEBSSESSSSSSSSSTIW 139
DB 121 PPGASOTIKSTTKRSPKPNKKTKKVISEEITE----- 156
QY 140 KIKSKNSANRELOKLLKVKDNKKNRTKKKPTKPPVVDAGSLDNGDFKVTPTST 199
DB 157 -----VKDNKKNRTKKKPTKPPVVDAGSLDNGDFKVTPTST 197
QY 200 TOHNKVSTSPKITTAKPINSRPSLPNSDTSKETSITVNETTETTKTQSTG 259
DB 198 TOHNKVSTSPKITTAKPINSRPSLPNSDTSKETSITVNETTETTKTQSTG 257
QY 260 KEKTSKAKESQSIKTSKADLAPTSKVLAKPTPAETTKGPAITTKETPTTKEPAS 319
DB 258 KEKTSKAKESQSIKTSKADLAPTSKVLAKPTPAETTKGPAITTKETPTTKEPAS 317
QY 320 TTPKEPTTIKSAPTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEP 379
DB 318 TTPKEPTTIKSAPTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEP 377
QY 380 APTTKSAPTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEP 439
DB 378 APTTKSAPTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEP 437
QY 440 EPAPTAPKKPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTT 499
DB 438 EPAPTAPKKPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTT 497
QY 500 TTKSAPTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTT 559
DB 498 TTKSAPTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTT 557
QY 560 APTAPKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEP 619
DB 558 APTAPKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEP 617
QY 620 PEEAPTTTAKAAAPNTKPEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTT 679
DB 618 PEEAPTTTAKAAAPNTKPEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTT 677
QY 680 APTTPKKAPKELAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTT 739
DB 678 APTTPKKAPKELAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTT 737
QY 740 TAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEP 799

Db 738 TAPTLKEPAPPTPKKPAKELAPPTTKGPTSTSDKPAPTTKETAPPTPKAPPTTPK 797
QY 800 KPAPPTPEPTPTTSVSTPTTKETTHKSPDESTPELSAEPKALENSPEKPGVPT 859
Db 798 KPAPPTPEPTPTTSVSTPTTKETTHKSPDESTPELSAEPKALENSPEKPGVPT 857
QY 860 TKTPAATKPEMTTAKDKTTERDLRTTPETTTAAAPKWKETATTTKTESKITATTQV 919
Db 858 TKTPAATKPEMTTAKDKTTERDLRTTPETTTAAAPKWKETATTTKTESKITATTQV 917
QY 920 TSTTTQDTPFKITLKITTLAPKVTTKKTTTTHIMNKPBEAKPKDRATNSKATTPK 979
Db 918 TSTTTQDTPFKITLKITTLAPKVTTKKTTTTHIMNKPBEAKPKDRATNSKATTPK 977
QY 980 POKPTKAPKPTSTKPKTMPRVRKPTTPTRKMTSTWPELNPTSRIAEAMLOTTTRPN 1039
Db 978 POKPTKAPKPTSTKPKTMPRVRKPTTPTRKMTSTWPELNPTSRIAEAMLOTTTRPN 1037
QY 1040 QTPNSKLVEVNPKSEDAGAGETHMLLRPHVFMPEVTPDMXYLPRVFNQIINPMLS 1099
Db 1038 QTPNSKLVEVNPKSEDAGAGETHMLLRPHVFMPEVTPDMXYLPRVFNQIINPMLS 1097

RESULT 11
US-07-757-022B-58
; Sequence 58, Application US/07757022B
; Patent No. 6433142
; GENERAL INFORMATION:
; APPLICANT: Gesner, Thomas G.
; APPLICANT: Clark, Stephen C.
; APPLICANT: Turner, Katherine
; APPLICANT: Hewick, Rodney M.
; TITLE OF INVENTION: Megakaryocyte Stimulating Factors
; NUMBER OF SEQUENCES: 143
; CORRESPONDENCE ADDRESS:
; ADDRESSES: Genetics Institute, Inc.
; STREET: 87 Cambridgepark Drive
; CITY: Cambridge
; STATE: Massachusetts
; COUNTRY: U.S.A.
; ZIP: 02140
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/757,022B
; FILING DATE: 19910910
; CLASSIFICATION: 530
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/643,502
; FILING DATE: 18-JAN-1991
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/546,114
; FILING DATE: 29-JUN-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/457,196
; FILING DATE: 29-DEC-1989
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/390,901
; FILING DATE: 08-AUG-1989
; ATTORNEY/AGENT INFORMATION:
; NAME: Cseri, Luann
; REGISTRATION NUMBER: 31,822
; REFERENCE/DOCKET NUMBER: GI 5190
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617)876-1170
; TELEFAX: (617)876-5851
; INFORMATION FOR SEQ ID NO: 58:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1049 amino acids

; TYPE: AMINO ACID
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-07-757-022B-58
Query Match 95.3%; Score 5561; DB 4; Length 1049;
Best Local Similarity 95.5%; Pred. No. 7.1e-162;
Matches 1049; Conservative 0; Mismatches 0; Indels 50; Gaps 1;
QY 1 MAWKTLPIYLLLLSVFVIOQVSSQBELSKGRGCFESFERGECDDCAQCKYDKCCPDYE 60
Db 1 MAWKTLPIYLLLLSVFVIOQVSSQBELSKGRGCFESFERGECDDCAQCKYDKCCPDYE 60
QY 61 SFCAEVENPTSPSSKKAPPPSCASQTIKSTTKRSKPPNKKTKKVISEETEBHSVS 120
Db 61 SFCA-----EHSVS 70
QY 121 ENQESSSSSSSSSTIWKIKSSKNSAANRELQKLVKDNKNQRTKKKPTPKPPVDE 180
Db 71 ENQESSSSSSSSSTIWKIKSSKNSAANRELQKLVKDNKNQRTKKKPTPKPPVDE 130
QY 181 AGSGLDNGDFKVTTPDTSTTOHNVKSTSPKITTAKFENRPSLPNNSDTSKETSLTWKE 240
Db 131 AGSGLDNGDFKVTTPDTSTTOHNVKSTSPKITTAKFENRPSLPNNSDTSKETSLTWKE 190
QY 241 TVVETKETTNNKQSTDGKEKITSASAKTSQIEKTSKADLAPTSKVLAKPTKAEITTKG 300
Db 191 TVVETKETTNNKQSTDGKEKITSASAKTSQIEKTSKADLAPTSKVLAKPTKAEITTKG 250
QY 301 PALTTKEPTPTTPKEPASTTKEPTTTIKSAPTTKEPAPTTTTSAPTTKEPAPTTT 360
Db 251 PALTTKEPTPTTPKEPASTTKEPTTTIKSAPTTKEPAPTTTTSAPTTKEPAPTTT 310
QY 361 KEPAPTTPKEPAPTTTKEPAPTTTTSAPTTKEPAPTTTKEPAPTTTKEPAPTTT 420
Db 311 KEPAPTTPKEPAPTTTKEPAPTTTTSAPTTKEPAPTTTKEPAPTTTKEPAPTTT 370
QY 421 TTPKEPAPTTKEPAPTTTKEPAPTTAPKAPPTTKEPAPTTTKEPAPTTTKEPAPTT 480
Db 371 TTPKEPAPTTKEPAPTTTKEPAPTTAPKAPPTTKEPAPTTTKEPAPTTTKEPAPTT 430
QY 481 PAPTTSAPTTTKEPAPTTTTSAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTT 540
Db 431 PAPTTSAPTTTKEPAPTTTTSAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTT 490
QY 541 KEPAPTTPKEPAPTTTKEPAPTTAPKAPPTTKEPAPTTTKEPAPTTTKEPAPTT 600
Db 491 KEPAPTTPKEPAPTTTKEPAPTTAPKAPPTTKEPAPTTTKEPAPTTTKEPAPTT 550
QY 601 PTTPELAPTTPEEPTPTTPKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTT 660
Db 551 PTTPELAPTTPEEPTPTTPKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTT 610
QY 661 KETAPTPKGTAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTT 720
Db 611 KETAPTPKGTAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTT 670
QY 721 KEPAPTTPKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTT 780
Db 671 KEPAPTTPKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTT 730
QY 781 TPKEPTAPTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTT 840
Db 731 TPKEPTAPTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTT 790
QY 841 APTPKALENSPEKPGVPTTKTTPAATKPEMTTAKDKTTERDLRTTPETTTAAAPKWKET 900
Db 791 APTPKALENSPEKPGVPTTKTTPAATKPEMTTAKDKTTERDLRTTPETTTAAAPKWKET 850
QY 901 ATTTEKTTESKITATTQVSTTTQDTPFKITLKITTLAPKVTTKKTTTTHIMNKP 960
Db 851 ATTTEKTTESKITATTQVSTTTQDTPFKITLKITTLAPKVTTKKTTTTHIMNKP 910

RESULT 13
 US-07-757-022B-48
 ; Sequence 48, Application US/07757022B
 ; Patent No. 6433142
 ; GENERAL INFORMATION:
 ; APPLICANT: Gesner, Thomas G.
 ; APPLICANT: Clark, Stephen C.
 ; APPLICANT: Turner, Katherine
 ; APPLICANT: Hewick, Rodney M.
 ; TITLE OF INVENTION: Megakaryocyte Stimulating Factors
 ; NUMBER OF SEQUENCES: 143
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Genetics Institute, Inc.
 ; STREET: 87 CambridgePark Drive
 ; CITY: Cambridge
 ; STATE: Massachusetts
 ; COUNTRY: U.S.A.
 ; ZIP: 02140
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: Patent in Release #1.0, Version #1.25
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/07/757,022B
 ; FILING DATE: 19910910
 ; CLASSIFICATION: 530
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: US 07/643,502
 ; FILING DATE: 18-JAN-1991
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: US 07/546,114
 ; FILING DATE: 29-JUN-1990
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: US 07/457,196
 ; FILING DATE: 29-DEC-1989
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: US 07/390,901
 ; FILING DATE: 08-AUG-1989
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Cseri, Luann
 ; REGISTRATION NUMBER: 31,822
 ; REFERENCE/DOCKET NUMBER: GI 5190
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: (617)876-1170
 ; TELEFAX: (617)876-5851
 ; INFORMATION FOR SEQ ID NO: 48:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 1354 amino acids
 ; TYPE: AMINO ACID
 ; TOPOLOGY: linear
 ; MOLECULE TYPE: protein
 ; US-07-757-022B-48

Query Match 95.1%; Score 5546.9; DB 4; Length 1354;
 Best Local Similarity 92.0%; Pred. No. 2.6e-161;
 Matches 1049; Conservative 0; Mismatches 0; Indels 91; Gaps 2;
 QY 1 MAWKTLPIYLLLSVFIQVSSQ----- 25
 DB 1 MAWKTLPIYLLLSVFIQVSSQDLSACAGCGEYSRATCNCYDNCQHYWCCPDF 60
 QY 26 -----ELSCKGRCFESFERGECDDAOCKYDKCCPDYESFCAEVHNPTSPSSKAP 79
 DB 61 KRVCTAELSCKGRCFESFERGECDDAOCKYDKCCPDYESFCA----- 105
 QY 90 PPGASQTIKSTTKRSPKPNKXKTKVISEEITEEHSVSENOESSSSSSSSSTIW 139
 DB 106 -----EEHSVSENOESSSSSSSSSSSTIW 130
 QY 140 KIKSSKNSAANRELQKLVKNDKNKNTKKKPTPKPPVWDEAGSLDNGDFKVTTPDTST 199

Db 131 KIKSSKNSAANRELQKLVKNDKNKNTKKKPTPKPPVWDEAGSLDNGDFKVTTPDTST 190
 QY 200 TQHNKYSTSPKITTAKPINRPSLPPNSDTSKETSJVNKETTIVETKETTNTKOTSTDG 259
 Db 191 TQHNKYSTSPKITTAKPINRPSLPPNSDTSKETSJVNKETTIVETKETTNTKOTSTDG 250
 QY 260 KEKTTSAKTSQTSKTSKDLAPTSKVLAKPTPKAETTTKGPALATTPKEPTTTKEPAS 319
 Db 251 KEKTTSAKTSQTSKTSKDLAPTSKVLAKPTPKAETTTKGPALATTPKEPTTTKEPAS 310
 QY 320 TTPKEPTTTIKSAPTTKPEAPTTTKSAPTTKPEAPTTTKPEAPTTTKPEAPTTTKPE 379
 Db 311 TTPKEPTTTIKSAPTTKPEAPTTTKSAPTTKPEAPTTTKPEAPTTTKPEAPTTTKPE 370
 QY 380 APTTKSAPTTKPEAPTTPKKAPTTKPEAPTTKPEAPTTKPEPTTPKPEAPTTKPEAPTT 439
 Db 371 APTTKSAPTTKPEAPTTPKKAPTTKPEAPTTKPEAPTTKPEPTTPKPEAPTTKPEAPTT 430
 QY 440 EPAPTAPKKAPTTKPEAPTTKPEAPTTKPEAPTTKPEAPTTTKSAPTTTKSAPTTTKPEAPT 499
 Db 431 EPAPTAPKKAPTTKPEAPTTKPEAPTTKPEAPTTTKERSPTTKPEAPTTTKSAPTTTKPEAPT 490
 QY 500 TTKSAPTTKERSPTTKPEAPTTKPEAPTTKPEAPTTKPEAPTTKPEAPTTTKKAP 559
 Db 491 TTKSAPTTKERSPTTKPEAPTTKPEAPTTKPEAPTTKPEAPTTKPEAPTTTKKAP 550
 QY 560 APTAPKPEAPTTPKETAPTTPKKLTPPTPEKLAPTTPEKPAFTTPEELAPTTPEEPTPTT 619
 Db 551 APTAPKPEAPTTPKETAPTTPKKLTPPTPEKLAPTTPEKPAFTTPEELAPTTPEEPTPTT 610
 QY 620 PEEPAPTTPKAAAPNTKPEAPTTKPEAPTTKPEAPTTKPEAPTTPKETAPTTPKGAPTTKPE 679
 Db 611 PEEPAPTTPKAAAPNTKPEAPTTKPEAPTTKPEAPTTKPEAPTTPKETAPTTPKGAPTTKPE 670
 QY 680 APTTKKAPKELAPTTTKETPTSTSDKPAFTTKGTAPTTKPEAPTTKPEAPTTKPEAPTTK 739
 Db 671 APTTKKAPKELAPTTTKETPTSTSDKPAFTTKGTAPTTKPEAPTTKPEAPTTKPEAPTTK 730
 QY 740 TAPTTKPEAPTTPKKAPKELAPTTTKGPTSTTSKPAFTTKETAPTTKPEAPTTKPEAPTTK 799
 Db 731 TAPTTKPEAPTTPKKAPKELAPTTTKGPTSTTSKPAFTTKETAPTTKPEAPTTKPEAPTTK 790
 QY 800 KPAPTTPEPTTSEVSTPTTKETPTTHKSPDESTPELSAETPKALENSKPEPGVPT 859
 Db 791 KPAPTTPEPTTSEVSTPTTKETPTTHKSPDESTPELSAETPKALENSKPEPGVPT 850
 QY 860 TKTPAATKPEMTTAKDKTTERDLATTPETTTAAPKMTKETATTTKTESKITATTTQV 919
 Db 851 TKTPAATKPEMTTAKDKTTERDLATTPETTTAAPKMTKETATTTKTESKITATTTQV 910
 QY 920 TSTTTQDTPPKITTLKTTTLAPKVTTKKIITTEIMNKSEETAKPKDRATNSKATTPK 979
 Db 911 TSTTTQDTPPKITTLKTTTLAPKVTTKKIITTEIMNKSEETAKPKDRATNSKATTPK 970
 QY 980 POKPTKAPKPTSTKKPKTMRVRKPTTTPPKMTSTMPKNTSRLAEAMLOTTTRPN 1039
 Db 971 POKPTKAPKPTSTKKPKTMRVRKPTTTPPKMTSTMPKNTSRLAEAMLOTTTRPN 1030
 QY 1040 QTPNSKLVENPKSEADAGAGETPHMLLRPHVFMPEVTPDMYDLPRVFNQGIINPMLS 1099
 Db 1031 QTPNSKLVENPKSEADAGAGETPHMLLRPHVFMPEVTPDMYDLPRVFNQGIINPMLS 1090

RESULT 14
 US-07-757-022B-50
 ; Sequence 50, Application US/07757022B
 ; Patent No. 6433142
 ; GENERAL INFORMATION:
 ; APPLICANT: Gesner, Thomas G.
 ; APPLICANT: Clark, Stephen C.
 ; APPLICANT: Turner, Katherine
 ; APPLICANT: Hewick, Rodney M.

TITLE OF INVENTION: Megakaryocyte Stimulating Factors
NUMBER OF SEQUENCES: 143
CORRESPONDENCE ADDRESS:
ADDRESSEE: Genetics Institute, Inc.
STREET: 87 Cambridgepark Drive
CITY: Cambridge
STATE: Massachusetts
COUNTRY: U.S.A.
ZIP: 02140
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US 07/57,022B
FILING DATE: 19910910
CLASSIFICATION: 530
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/643,502
FILING DATE: 18-JAN-1991
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/546,114
FILING DATE: 29-JUN-1990
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/457,196
FILING DATE: 29-DEC-1989
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/390,901
FILING DATE: 08-AUG-1989
ATTORNEY/AGENT INFORMATION:
NAME: Cserr, Luann
REGISTRATION NUMBER: 31,922
REFERENCE/DOCKET NUMBER: GI 5190
TELEPHONE: (617) 876-1170
TELEFAX: (617) 876-5851
INFORMATION FOR SEQ ID NO: 50:
SEQUENCE CHARACTERISTICS:
LENGTH: 1314 amino acids
TYPE: AMINO ACID
TOPOLOGY: linear
MOLECULE TYPE: protein
US-07-757-022B-50

Query Match 93.2%; Score 5437.9; DB 4; Length 1314;
Best Local Similarity 93.5%; Pred. No. 5.3e-158;
Matches 1028; Conservative 7; Mismatches 14; Indels 51; Gaps 2;
QY 1 MAWKTLPIYLLLLSVFVIQQVSSQEL-SCKGRCESEPERGECCDAQCKYDKCCPDY 59
DB 1 MAWKTLPIYLLLLSVFVIQQVSSQELSSCAGRGEGYSDATCNCIDYNCQHYNECCPDF 60
QY 60 ESFCAEVHNPSTGPPSSKKAPPPSGASQTIKSTTKRSPKPPNKKTKKVIESEBITEHSV 119
DB 61 KRVG-----TAEHSV 70
QY 120 SENQSSSSSSSSSSSTTWIKSSKNSAANRELQKLVKONKKNRTKKKTPKPPVD 179
DB 71 SENQSSSSSSSSSSSTTWIKSSKNSAANRELQKLVKONKKNRTKKKTPKPPVD 130
QY 180 EAGSGLDNGDFKVTTPDTSTQHNKYSTSPKITTAKPINRPSLPNSDTSKETSITVVK 239
DB 131 EAGSGLDNGDFKVTTPDTSTQHNKYSTSPKITTAKPINRPSLPNSDTSKETSITVVK 190
QY 240 ETVETKETTNNKQTSQCKEKTSAKETQSIKTSKADLAPTSKVLAKPTPKAETTK 299
DB 191 ETVETKETTNNKQTSQCKEKTSAKETQSIKTSKADLAPTSKVLAKPTPKAETTK 250
QY 300 GPALTTPKPTTPPKPEASTTPKEPTPTTKSAPTTPKPEAPTTTKSAPTTPKPEAPT 359
DB 251 GPALTTPKPTTPPKPEASTTPKEPTPTTKSAPTTPKPEAPTTTKSAPTTPKPEAPT 310

QY 360 TKPEAPTTKEPAPTTTKPEAPTTTKSAPTTTKPEAPTTTKPEAPTTTKPEAPTTKEPT 419
DB 311 TKPEAPTTKEPAPTTTKPEAPTTTKSAPTTTKPEAPTTTKPEAPTTTKPEAPTTKEPT 370
QY 420 PTTKPSAPTTKEPAPTTTKPEAPTTAPKAPAPTTTKPEAPTTTKPEAPTTTKPSPTTK 479
DB 371 PTTKPSAPTTKEPAPTTTKPEAPTTAPKAPAPTTTKPEAPTTTKPEAPTTTKPSPTTK 430
QY 480 EPAPTTTKSAPTTTKPEAPTTTKSAPTTTKPEAPTTTKPEAPTTTKPEAPTTTKPEAPT 539
DB 431 EPAPTTTKSAPTTTKPEAPTTTKSAPTTTKPEAPTTTKPEAPTTTKPEAPTTTKPEAPT 490
QY 540 PKPSAPTTKEPAPTTTKKAPAPTTAPKAPAPTTTKPSAPTTTKPEAPTTTKPEAPTTKEP 599
DB 491 PKPSAPTTKEPAPTTTKKAPAPTTAPKAPAPTTTKPSAPTTTKPEAPTTTKPEAPTTKEP 550
QY 600 APTEBELAPTTPEEPTTPPEAPTTPKAAAPNTPKPEAPTTPKPEAPTTTKPEAPTT 659
DB 551 APTEBELAPTTPEEPTTPPEAPTTPKAAAPNTPKPEAPTTPKPEAPTTTKPEAPTT 610
QY 660 PKETAPTTPKGTAPTTLKEPAPTTPKKAPAPKELAPTTTKETSTTSKAPAPTTKGTAPT 719
DB 611 PKETAPTTPKGTAPTTLKEPAPTTPKKAPAPKELAPTTTKETSTTSKAPAPTTKGTAPT 670
QY 720 TKPEAPTTKEPAPTTTKGTAPTTLKEPAPTTPKKAPAPKELAPTTTKGTSTTSKAPAP 779
DB 671 TKPEAPTTKEPAPTTTKGTAPTTLKEPAPTTPKKAPAPKELAPTTTKGTSTTSKAPAP 730
QY 780 TTPKETAPTTKEPAPTTPKKAPAPTTPEPTTPPTTSEVSTTTTKEPTTIHKSPDESTPEL 839
DB 731 TTPKETAPTTKEPAPTTPKKAPAPTTPEPTTPPTTSEVSTTTTKEPTTIHKSPDESTPEL 790
QY 840 SAETPKALENSKPEGVPTTKPAATKPEMTTAKDKITERDLRTTETTTAAKMTKE 899
DB 791 SAETPKALENSKPEGVPTTKPAATKPEMTTAKDKITERDLRTTETTTAAKMTKE 850
QY 900 TATTEKTESKITATTQVSTTTQDTPFKITLTKTTTLAPKVITTKITITTEIMNK 959
DB 851 TATTEKTESKITATTQVSTTTQDTPFKITLTKTTTLAPKVITTKITITTEIMNK 910
QY 960 PEETAKPKDRATNSKATTPKQPKTKAPKPTSTTKPKTWPVRVKPTTTPRKWTSTMP 1019
DB 911 PEETAKPKDRATNSKATTPKQPKTKAPKPTSTTKPKTWPVRVKPTTTPRKWTSTMP 970
QY 1020 ELNPTSRIAEAMLOTTTRPNQTNPSKLVEYNPKSEDAGGAEGETPHMLLRPHVFMPEVTP 1079
DB 971 ELNPTSRIAEAMLOTTTRPNQTNPSKLVEYNPKSEDAGGAEGETPHMLLRPHVFMPEVTP 1030
QY 1080 DMDYLPRVNPQGIINPMLS 1099
DB 1031 DMDYLPRVNPQGIINPMLS 1050
RESULT 15
US-07-757-022B-74
; Sequence 74, Application US/07757022B
; Patent No. 6433142
; GENERAL INFORMATION:
; APPLICANT: Gesner, Thomas G.
; APPLICANT: Clark, Stephen C.
; APPLICANT: Turner, Katherine
; APPLICANT: Hewick, Rodney M.
; TITLE OF INVENTION: Megakaryocyte Stimulating Factors
; NUMBER OF SEQUENCES: 143
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genetics Institute, Inc.
; STREET: 87 Cambridgepark Drive
; CITY: Cambridge
; STATE: Massachusetts
; COUNTRY: U.S.A.
; ZIP: 02140
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk

```

; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/757,022B
; FILING DATE: 19910910
; CLASSIFICATION: 530
; PRIORITY DATA:
; APPLICATION NUMBER: US 07/643,502
; FILING DATE: 18-JAN-1991
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/546,114
; FILING DATE: 29-JUN-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/457,196
; FILING DATE: 29-DEC-1989
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/390,901
; FILING DATE: 08-AUG-1989
; ATTORNEY/AGENT INFORMATION:
; NAME: Cserny, Luann
; REGISTRATION NUMBER: 31,822
; REFERENCE/DOCKET NUMBER: GI 5190
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617)876-1170
; TELEFAX: (617)876-5851
; INFORMATION FOR SEQ ID NO: 74:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1038 amino acids
; TYPE: AMINO ACID
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-07-757-022B-74

```

| | | | | | | | | |
|-----------------------|--------------|------------|------------------|----------------------------|-----------------------|---------------|-------|----|
| Query Match | 91.8%; | Score | 5353.7; | DB | 4; | Length | 1038; | |
| Best Local Similarity | 91.5%; | Pred. No. | 1.4e-155; | | | | | |
| Matches 1006; | Conservative | 0; | Mismatches | 0; | Indels | 93; | Gaps | 1; |
| Qy | 1 | MAWKTLPIYL | LLLLLSVFVI | QOVSSQELSSCKGRCFESF | RGREDCDDAOCKKYDKCCPYE | 60 | | |
| Db | 1 | MAWKTLPIYL | LLLLLSVFVI | QOVSSQELSSCKGRCFESF | RGREDCDDAOCKKYDKCCPYE | 60 | | |
| Qy | 61 | SFCAEVHNPT | SPPSKKAPP | SGASQTIKSTTKRSPKPKKKKTKKVI | TESEBITTEHSVS | 120 | | |
| Db | 61 | SFCAB | ----- | ----- | ----- | 65 | | |
| Qy | 121 | ENQESSSSSS | SSSSSTIWKIKSKKSN | SAANRELQKKLVKONKORTKKKPT | PKPPVVDE | 180 | | |
| Db | 66 | ----- | ----- | ----- | ----- | 87 | | |
| Qy | 181 | AGSLDNGDF | KVTPTDTS | TTQHNKVSTSPKITTAKPINRPSL | PNSDTSKETS | SLTWKE | 240 | |
| Db | 88 | AGSLDNGDF | KVTPTDTS | TTQHNKVSTSPKITTAKPINRPSL | PNSDTSKETS | SLTWKE | 147 | |
| Qy | 241 | TTVETKET | TTTNKQTS | TDGKEKTTSAKETQS | IEKTSADLAPTSKVLAKPT | PKAETTTKG | 300 | |
| Db | 148 | TTVETKET | TTTNKQTS | TDGKEKTTSAKETQS | IEKTSADLAPTSKVLAKPT | PKAETTTKG | 207 | |
| Qy | 301 | PALTTPK | EPPTTTPK | EPASTPKEPPTTIK | GAPTTPKEPAPTTTKSAPTT | TKESPAPTTT | 360 | |
| Db | 208 | PALTTPK | EPPTTTPK | EPASTPKEPPTTIK | GAPTTPKEPAPTTTKSAPTT | TKESPAPTTT | 267 | |
| Qy | 361 | KEPAPTT | PKEPAPTT | TKEPAPTTTKSAPTT | PKEPAPTTTKKAPTT | TKEPAPTTKESPT | 420 | |
| Db | 268 | KEPAPTT | PKEPAPTT | TKEPAPTTTKGAPTT | PKEPAPTTPKKAPTT | TKEPAPTTKESPT | 327 | |
| Qy | 421 | TTPK | EPAPTTKEPAPTT | PKEPAPTA | PKKAPTTTPKEPAPTTT | PKESPTTPKE | 480 | |
| Db | 328 | TTPK | EPAPTTKEPAPTT | PKEPAPTA | PKKAPTTTPKEPAPTTT | PKESPTTPKE | 387 | |
| Qy | 481 | PAPTTTK | SAPTTTK | EPAPTTTKSAPTT | PKESPPTTTKGPAPTT | TKGPAPTTT | 540 | |
| Db | 388 | PAPTTTK | SAPTTTK | EPAPTTTKSAPTT | PKESPPTTTKGPAPTT | TKGPAPTTT | 447 | |

| | | | |
|----|------|---|------|
| Qy | 541 | KEPAPTTPKGPAPTTTKKAPAPTAKEBPAPTTPKBTAPTTPKLPTTPEKLAPTTPKPA | 600 |
| Db | 448 | KEPAPTTPKGPAPTTTKKAPAPTAKEBPAPTTPKBTAPTTPKLPTTPEKLAPTTPKPA | 507 |
| Qy | 601 | PTTPEELAPTTPEBPPTTPEEPAPTTPKAAAPNTKEPAPTTPKBPAPTTPKBPAPTTT | 660 |
| Db | 508 | PTTPEELAPTTPEBPPTTPEEPAPTTPKAAAPNTKEPAPTTPKBPAPTTPKBPAPTTT | 567 |
| Qy | 661 | KETAPTTPKGTAPTTLKEBPAPTTPKKAPKELAPTTTKETSTTSDKPAPTTPKGTAPTT | 720 |
| Db | 568 | KETAPTTPKGTAPTTLKEBPAPTTPKKAPKELAPTTTKETSTTSDKPAPTTPKGTAPTT | 627 |
| Qy | 721 | PKGPAPTTKEBPAPTTPKGTAPTTTLKEBPAPTTPKKAPKELAPTTTKGTSTTSDKPAPTT | 780 |
| Db | 628 | PKGPAPTTKEBPAPTTPKGTAPTTTLKEBPAPTTPKKAPKELAPTTTKGTSTTSDKPAPTT | 687 |
| Qy | 781 | TPKETAPTTKEBPAPTTPKBPAPTTBETPPPTTSEVSTPTTTKEPTTIHKSPDESTPELS | 840 |
| Db | 688 | TPKETAPTTKEBPAPTTPKBPAPTTBETPPPTTSEVSTPTTTKEPTTIHKSPDESTPELS | 747 |
| Qy | 841 | AETPKALENSPKEPGVPTTKTAAATKPEMTTTAKDKTTERDRLTTPETTTAAAPKMTKET | 900 |
| Db | 748 | AETPKALENSPKEPGVPTTKTAAATKPEMTTTAKDKTTERDRLTTPETTTAAAPKMTKET | 807 |
| Qy | 901 | ATTEKTESKITATTQVTSSTTQDTPPKITTLKTTLAPKVTTTKKITTTTEIMNKP | 960 |
| Db | 808 | ATTEKTESKITATTQVTSSTTQDTPPKITTLKTTLAPKVTTTKKITTTTEIMNKP | 867 |
| Qy | 961 | EETAKKORATNSKATTPKPKKTSKAPKPKTSKKKPKTMRVRKPKTTTTPRKWTSMTPE | 1020 |
| Db | 866 | EETAKKORATNSKATTPKPKKTSKAPKPKTSKKKPKTMRVRKPKTTTTPRKWTSMTPE | 927 |
| Qy | 1021 | LNPTSIAEAMLQTTTRPNQTPNSKLVEVNPKSEDAGGAEGETPHMLLRPHVFMPEVTPD | 1080 |
| Db | 928 | LNPTSIAEAMLQTTTRPNQTPNSKLVEVNPKSEDAGGAEGETPHMLLRPHVFMPEVTPD | 987 |
| Qy | 1081 | MDYLPRVPNOGIIINPMLS 1099 | |
| Db | 988 | MDYLPRVPNOGIIINPMLS 1006 | |

Search completed: October 13, 2004, 11:58:45
Job time : 25.1039 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model
Run on: October 13, 2004, 11:23:49 ; Search time 111.81 Seconds
(without alignments)
3171.696 Million cell updates/sec

Title: SEQ1-A
Perfect score: 5835
Sequence: 1 MAWKTLPIYLLLSVFVIQ.....DMDYLPRVNGIINPMLS 1099

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.1

Searched: 1356558 seqs, 322692953 residues

Total number of hits satisfying chosen parameters: 1356558

Minimum DB seq length: 0
Maximum DB seq length: 200000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

- Database : Published Applications AA.*
- 1: /cgn2_6/ptodata/1/pubpaa/US07_PUBCOMB.pep.*
 - 2: /cgn2_6/ptodata/1/pubpaa/PCT_NEW_PUB.pep.*
 - 3: /cgn2_6/ptodata/1/pubpaa/US06_NEW_PUB.pep.*
 - 4: /cgn2_6/ptodata/1/pubpaa/US06_PUBCOMB.pep.*
 - 5: /cgn2_6/ptodata/1/pubpaa/US07_NEW_PUB.pep.*
 - 6: /cgn2_6/ptodata/1/pubpaa/PCTUS_PUBCOMB.pep.*
 - 7: /cgn2_6/ptodata/1/pubpaa/US08_NEW_PUB.pep.*
 - 8: /cgn2_6/ptodata/1/pubpaa/US08_PUBCOMB.pep.*
 - 9: /cgn2_6/ptodata/1/pubpaa/US09A_PUBCOMB.pep.*
 - 10: /cgn2_6/ptodata/1/pubpaa/US09B_PUBCOMB.pep.*
 - 11: /cgn2_6/ptodata/1/pubpaa/US09C_PUBCOMB.pep.*
 - 12: /cgn2_6/ptodata/1/pubpaa/US09_NEW_PUB.pep.*
 - 13: /cgn2_6/ptodata/1/pubpaa/US10A_PUBCOMB.pep.*
 - 14: /cgn2_6/ptodata/1/pubpaa/US10B_PUBCOMB.pep.*
 - 15: /cgn2_6/ptodata/1/pubpaa/US10C_PUBCOMB.pep.*
 - 16: /cgn2_6/ptodata/1/pubpaa/US10D_PUBCOMB.pep.*
 - 17: /cgn2_6/ptodata/1/pubpaa/US10_NEW_PUB.pep.*
 - 18: /cgn2_6/ptodata/1/pubpaa/US60_NEW_PUB.pep.*
 - 19: /cgn2_6/ptodata/1/pubpaa/US60_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | ID | Description |
|------------|--------|-------------|--------|----|-------------------|
| 1 | 5835 | 100.0 | 1363 | 13 | US-10-124-557-52 |
| 2 | 5820.9 | 99.8 | 1140 | 13 | US-10-124-557-104 |
| 3 | 5820.9 | 99.8 | 1404 | 9 | US-09-802-207-30 |
| 4 | 5820.9 | 99.8 | 1404 | 11 | US-09-897-188-1 |
| 5 | 5820.9 | 99.8 | 1404 | 13 | US-10-124-557-2 |
| 6 | 5820.9 | 99.8 | 1404 | 13 | US-10-124-557-62 |
| 7 | 5817.7 | 96.3 | 1320 | 13 | US-10-124-557-46 |
| 8 | 5817.7 | 96.3 | 1320 | 13 | US-10-124-557-60 |
| 9 | 5803.6 | 96.0 | 1361 | 13 | US-10-124-557-40 |
| 10 | 5561 | 95.3 | 1049 | 13 | US-10-124-557-58 |
| 11 | 5561 | 95.1 | 1313 | 13 | US-10-124-557-142 |
| 12 | 5846.9 | 95.1 | 1354 | 13 | US-10-124-557-48 |
| 13 | 5837.9 | 93.2 | 1314 | 13 | US-10-124-557-50 |
| 14 | 5353.7 | 91.8 | 1038 | 13 | US-10-124-557-74 |

- 15 5353.7 91.8 1270 13 US-10-124-557-44
- 16 5339.6 91.5 1311 13 US-10-124-557-42
- 17 5331.9 89.7 1022 13 US-10-124-557-84
- 18 5011 85.9 941 13 US-10-124-557-14
- 19 2456.7 42.1 732 9 US-09-802-207-27
- 20 1409.9 24.2 538 14 US-10-038-694-3
- 21 1401.6 24.0 292 16 US-10-468-910-4
- 22 1354.4 23.2 5179 9 US-09-922-217-1068
- 23 1354.4 23.2 5179 9 US-09-833-263-1068
- 24 1354.4 23.2 5179 9 US-10-025-380-1068
- 25 1354.4 23.2 5179 16 US-10-734-564-121
- 26 972.9 16.7 1460 14 US-10-295-027-428
- 27 953.6 16.3 1325 9 US-09-864-761-35812
- 28 933.8 16.0 1367 9 US-09-801-368-108
- 29 923.9 15.8 220 13 US-10-124-557-96
- 30 896.2 15.4 3507 14 US-10-369-493-5784
- 31 879.9 15.1 6642 14 US-10-369-493-5013
- 32 877.6 15.0 2090 16 US-10-408-765A-2318
- 33 864.6 14.8 207 13 US-10-124-557-116
- 34 864.6 14.8 207 13 US-10-124-557-136
- 35 863.5 14.8 3256 16 US-10-408-765A-174
- 36 862.2 14.8 422 13 US-10-701-490-9
- 37 861.9 14.8 209 13 US-10-124-557-68
- 38 860.7 14.8 1151 10 US-10-124-557-94
- 39 850.6 14.7 185 16 US-09-825-751A-79
- 40 850.6 14.7 208 13 US-10-468-910-2
- 41 856.9 14.7 22152 16 US-10-124-557-132
- 42 856.7 14.7 5935 14 US-10-715-066-5
- 43 855.8 14.7 3256 9 US-10-243-243A-8
- 44 855.7 14.7 3256 9 US-09-919-172-98
- 45 855.7 14.7 3256 10 US-09-919-039-21

ALIGNMENTS

RESULT 1
US-10-124-557-52
; Sequence 52 Application US/10124557
; Publication No. US20020137894A1
; GENERAL INFORMATION:
; APPLICANT: Turner, Katherine
; Clark, Stephen C.
; Jacobs, Kenneth
; Hewick, Rodney M.
; Gesner, Thomas G.
; TITLE OF INVENTION: Megakaryocyte Stimulating Factors
; NUMBER OF SEQUENCES: 143
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genetics Institute, Inc.
; STREET: 87 CambridgePark Drive
; CITY: Cambridge
; STATE: Massachusetts
; COUNTRY: U.S.A.
; ZIP: 02140
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/10/124,557
; FILING DATE: 16-Apr-2002
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/643,502
; FILING DATE: 18-JAN-1991
; APPLICATION NUMBER: US 07/546,114
; FILING DATE: 29-JUN-1990
; APPLICATION NUMBER: US 07/457,196
; FILING DATE: 29-DEC-1989
; APPLICATION NUMBER: US 07/390,901
; FILING DATE: 08-AUG-1989

```

;
; ATTORNEY/AGENT INFORMATION:
; NAME: Cseri, Luann
; REGISTRATION NUMBER: 31,822
; REFERENCE/DOCKET NUMBER: GI 5190
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617)876-1170
; TELEFAX: (617)876-5851
; INFORMATION FOR SEQ ID NO: 52:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1363 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; SEQUENCE DESCRIPTION: SEQ ID NO: 52:
US-10-124-557-52

Query Match 100.0%; Score 5835; DB 13; Length 1363;
Best Local Similarity 100.0%; Pred. No. 2.8e-142;
Matches 1099; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MAWKTLPIYLLLLSVFVLIQVSSOELSCGRCFSPERGRECDCAQCKYDKCCPDYE 60
Db 1 MAWKTLPIYLLLLSVFVLIQVSSOELSCGRCFSPERGRECDCAQCKYDKCCPDYE 60

QY 61 SFCAEVHNPSPSSKXAPPSPGASQTIKSTTKRSPKPPNKKTKKVIESEBIEEHSVS 120
Db 61 SFCAEVHNPSPSSKXAPPSPGASQTIKSTTKRSPKPPNKKTKKVIESEBIEEHSVS 120

QY 121 ENQESSSSSSSSSTTWIKSSKNSAANRELQKLVKONKQRTKKKTPKPPVDE 180
Db 121 ENQESSSSSSSSSTTWIKSSKNSAANRELQKLVKONKQRTKKKTPKPPVDE 180

QY 181 AGSGLNDGDKVTTTDTSTQHKVSTSGPKITAKPINRPSLPNSDTSKETSILTWNKE 240
Db 181 AGSGLNDGDKVTTTDTSTQHKVSTSGPKITAKPINRPSLPNSDTSKETSILTWNKE 240

QY 241 TVVETKETTNNKQSTDGKETTSAKETQSIKTSAXDLAPTSKVLAQPKPAETTKG 300
Db 241 TVVETKETTNNKQSTDGKETTSAKETQSIKTSAXDLAPTSKVLAQPKPAETTKG 300

QY 301 PALTTPKETPTTPKPEASTTPKEPTTIISAPTTKPEAPTTTKSAPTTKPEAPTTT 360
Db 301 PALTTPKETPTTPKPEASTTPKEPTTIISAPTTKPEAPTTTKSAPTTKPEAPTTT 360

QY 361 KEPAPTTKPEAPTTTKSAPTTTKSAPTTKPEAPTTPKKPAPTTPKPEAPTTPKSP 420
Db 361 KEPAPTTKPEAPTTTKSAPTTTKSAPTTKPEAPTTPKKPAPTTPKPEAPTTPKSP 420

QY 421 TTPKEPAPTTKPEAPTTKPEAPTAQKPAPTTPKPEAPTTKPEAPTTKPEAPTTKE 480
Db 421 TTPKEPAPTTKPEAPTTKPEAPTAQKPAPTTPKPEAPTTKPEAPTTKPEAPTTKE 480

QY 481 PAPITTKSAPTTKPEAPTTTKSAPTTTKPEAPTTKPEAPTTKPEAPTTPKKPA 540
Db 481 PAPITTKSAPTTKPEAPTTTKSAPTTTKPEAPTTKPEAPTTKPEAPTTPKKPA 540

QY 541 KEPAPTTKPEAPTTKPEAPTAQKPAPTTPKPEAPTTKPEAPTTKPEAPTTPKK 600
Db 541 KEPAPTTKPEAPTTKPEAPTAQKPAPTTPKPEAPTTKPEAPTTKPEAPTTPKK 600

QY 601 PTTPEELAPTTPEEPTPTTPPEAPTTPKAAAPNTKPEAPTTKPEAPTTKPEAPTT 660
Db 601 PTTPEELAPTTPEEPTPTTPPEAPTTPKAAAPNTKPEAPTTKPEAPTTKPEAPTT 660

QY 661 KETAPTTKGAFTAPTLKPEAPTTKPKAPKELAPTTTKGPTSTTSDKPAPTT 720
Db 661 KETAPTTKGAFTAPTLKPEAPTTKPKAPKELAPTTTKGPTSTTSDKPAPTT 720

QY 721 PKEAPTTKPEAPTTKGAFTAPTLKPEAPTTKPKAPKELAPTTTKGPTSTTSDK 780
Db 721 PKEAPTTKPEAPTTKGAFTAPTLKPEAPTTKPKAPKELAPTTTKGPTSTTSDK 780

QY 781 TPKETAPTTKPEAPTTKPKAPPTTPPTTSEVSTPTTKEPTTIHKSPDESTPELS 840

```

```

;
; Db 781 TPKETAPTTKPEAPTTKPKAPPTTPPTTSEVSTPTTKEPTTIHKSPDESTPELS 840
;
; QY 841 AETPKALENSPKPGVPTTKTAAATKPEWTTTAKOKTTERDLRTTPTTAAAPKMTKET 900
;
; Db 841 AETPKALENSPKPGVPTTKTAAATKPEWTTTAKOKTTERDLRTTPTTAAAPKMTKET 900
;
; QY 901 ATTEKTESKITATTQVTSITTQDTPFKITLTKTTLAPKVTTTKTITTEIMNKP 960
;
; Db 901 ATTEKTESKITATTQVTSITTQDTPFKITLTKTTLAPKVTTTKTITTEIMNKP 960
;
; QY 961 EETAKPKDRATNSKATTPKPKETKAPKAPTSTSKPKETMPVRVKPATTTPKMTSTME 1020
;
; Db 961 EETAKPKDRATNSKATTPKPKETKAPKAPTSTSKPKETMPVRVKPATTTPKMTSTME 1020
;
; QY 1021 LNPTSRIAEAMLOTTTRPNQTPNSKLVEVNPXSSEDAGGAEGTPHMLLRPHVFMPEVTPD 1080
;
; Db 1021 LNPTSRIAEAMLOTTTRPNQTPNSKLVEVNPXSSEDAGGAEGTPHMLLRPHVFMPEVTPD 1080
;
; QY 1081 MDYLPKVPNQGIIINPMLS 1099
;
; Db 1081 MDYLPKVPNQGIIINPMLS 1099
;
; RESULT 2
; US-10-124-557-104
; Sequence 104, Application US/10124557
; Publication No. US2002037894A1
; GENERAL INFORMATION:
; APPLICANT: Turner, Katherine
; Clark, Stephen C.
; Jacobs, Kenneth
; Hewick, Rodney M.
; Gesner, Thomas G.
; TITLE OF INVENTION: Megakaryocyte Stimulating Factors
; NUMBER OF SEQUENCES: 143
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genetics Institute, Inc.
; STREET: 87 CambridgePark Drive
; CITY: Cambridge
; STATE: Massachusetts
; COUNTRY: U.S.A.
; ZIP: 02140
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent in Release #1.0, Version #1.25
; CURRENT APPLICATION NUMBER: US/10/124,557
; FILING DATE: 16-Apr-2002
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/643,502
; FILING DATE: 18-JAN-1991
; APPLICATION NUMBER: US 07/546,114
; FILING DATE: 29-JUN-1990
; APPLICATION NUMBER: US 07/457,196
; FILING DATE: 29-DEC-1989
; APPLICATION NUMBER: US 07/390,901
; FILING DATE: 08-AUG-1989
; ATTORNEY/AGENT INFORMATION:
; NAME: Cseri, Luann
; REGISTRATION NUMBER: 31,822
; REFERENCE/DOCKET NUMBER: GI 5190
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617)876-1170
; TELEFAX: (617)876-5851
; INFORMATION FOR SEQ ID NO: 104:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1140 amino acids
; TYPE: amino acid
; TOPOLOGY: linear

```

```

; MOLECULE TYPE: protein
; SEQUENCE DESCRIPTION: SEQ ID NO: 104:
US-10-124-557-104

Query Match          99.8%; Score 5820.9; DB 13; Length 1140;
Best Local Similarity 96.4%; Pred. No. 5.2e-142;
Matches 1099; Conservative 0; Mismatches 0; Indels 41; Gaps 1;

QY 1 MAWKTLPIYLLLLSVFVIQVSSQ----- 25
DB 1 MAWKTLPIYLLLLSVFVIQVSSQDLSSCAGRGEGYSDATCNCYDNCQHYMECCPDF 60

QY 26 -----ELSCKGRCFESFERGECDDAQQCKYDKCCPDYEFCAEVHNPTSPSSSKAP 79
DB 61 KRVTAEELSCKGRCFESFERGECDDAQQCKYDKCCPDYEFCAEVHNPTSPSSSKAP 120

QY 80 PPSGASQTIKSTTKSPKPNKKTKVIESEIEEHSVSENQESSSSSSSSSTI 139
DB 121 PPSGASQTIKSTTKSPKPNKKTKVIESEIEEHSVSENQESSSSSSSSSTI 180

QY 140 KIKSSKNSAANRELQKLVKDKNKNRTKKKPTPKPPVVDGAGSLDNGDFKVTPTST 199
DB 181 KIKSSKNSAANRELQKLVKDKNKNRTKKKPTPKPPVVDGAGSLDNGDFKVTPTST 240

QY 200 TOHNKVSTSPKITTAKPINRPSLPNSDTSKETSLSLVNKKETVETKETTNNKQSTSDG 259
DB 241 TOHNKVSTSPKITTAKPINRPSLPNSDTSKETSLSLVNKKETVETKETTNNKQSTSDG 300

QY 260 KEKTTSAKETQSIKTSKADLAPTSKVLAKPTPKAETTTKGPALTTKPEPTTPKPEAS 319
DB 301 KEKTTSAKETQSIKTSKADLAPTSKVLAKPTPKAETTTKGPALTTKPEPTTPKPEAS 360

QY 320 TTPKEPTPTTIKSAPTTPKPEAPTTTKGAPTTKSPAPTTTKPEAPTTTKPEAPTTTKEP 379
DB 361 TTPKEPTPTTIKSAPTTPKPEAPTTTKGAPTTKSPAPTTTKPEAPTTTKPEAPTTTKEP 420

QY 380 APTTKSAPTTTKPEAPTTTKPEAPTTTKPEAPTTTKPEAPTTTKPEAPTTTKPEAPTPK 439
DB 421 APTTKSAPTTTKPEAPTTTKPEAPTTTKPEAPTTTKPEAPTTTKPEAPTTTKPEAPTPK 480

QY 440 EPAPTAPKKPAPTTTKPEAPTTTKPEAPTTTKPEAPTTTKPEAPTTTKSAPTTTKEPAPT 499
DB 481 EPAPTAPKKPAPTTTKPEAPTTTKPEAPTTTKPEAPTTTKPEAPTTTKSAPTTTKEPAPT 540

QY 500 TTKSAPTTTKPEAPTTTKPEAPTTTKPEAPTTTKPEAPTTTKPEAPTTTKPEAPTTTKKP 559
DB 541 TTKSAPTTTKPEAPTTTKPEAPTTTKPEAPTTTKPEAPTTTKPEAPTTTKPEAPTTTKKP 600

QY 560 APTAPKEAPTTTKPEAPTTTKPEAPTTTKPEAPTTTKPEAPTTTKPEAPTTTKPEAPTTT 619
DB 601 APTAPKEAPTTTKPEAPTTTKPEAPTTTKPEAPTTTKPEAPTTTKPEAPTTTKPEAPTTT 660

QY 620 PEEAPTTPKAAAPNTKPEAPTTTKPEAPTTTKPEAPTTTKPEAPTTTKPEAPTTTKPEAPTT 679
DB 661 PEEAPTTPKAAAPNTKPEAPTTTKPEAPTTTKPEAPTTTKPEAPTTTKPEAPTTTKPEAPTT 720

QY 680 APTTPKKAPKELAPTTTKPEPTSTSDKPAITTKGAPTTTKPEAPTTTKPEAPTTTKPG 739
DB 721 APTTPKKAPKELAPTTTKPEPTSTSDKPAITTKGAPTTTKPEAPTTTKPEAPTTTKPEAPTT 780

QY 740 TAPTTLKEPAPTTTKPEAPTTTKPEAPTTTKPEAPTTTKPEAPTTTKPEAPTTTKPEAPTT 799
DB 781 TAPTTLKEPAPTTTKPEAPTTTKPEAPTTTKPEAPTTTKPEAPTTTKPEAPTTTKPEAPTT 840

QY 800 KPAPTTPETPPPTTSVSTPTTKPTTIHKSPDESSTPELSAETPKALENSPKPGVPT 859
DB 841 KPAPTTPETPPPTTSVSTPTTKPTTIHKSPDESSTPELSAETPKALENSPKPGVPT 900

QY 860 TKTPAATKEMTTTAKDKTTERDLRTTPTTTTAAAPKMTKETATTTTEKTTESKITATTTQV 919
DB 901 TKTPAATKEMTTTAKDKTTERDLRTTPTTTTAAAPKMTKETATTTTEKTTESKITATTTQV 960

QY 920 TSTTTQDTPFPKITTLLKTTLLAPKVTTTKKTTTTTTEIMNKPBEETAKPKDRATNSKATTPK 979

```

```

DB 961 TSTTTQDTPFPKITTLLKTTLLAPKVTTTKKTTTTTTEIMNKPBEETAKPKDRATNSKATTPK 1020
QY 980 PQPPTKAPKPPSTTKKPKTMPVRKPKKTTTPEKMTSTMPBELNPTSRIAEAMLOTTTRN 1039
DB 1021 PQPPTKAPKPPSTTKKPKTMPVRKPKKTTTPEKMTSTMPBELNPTSRIAEAMLOTTTRN 1080
QY 1040 QTPNSKLIVEVNPKSEDAAGAGETPHMLLRPHVFMPEVTPDMDYLPRVFNQGIINPMLS 1099
DB 1081 QTPNSKLIVEVNPKSEDAAGAGETPHMLLRPHVFMPEVTPDMDYLPRVFNQGIINPMLS 1140

RESULT 3
US-09-802-207-30
; Sequence 30, Application US/09802207
; Publication No. US20020086824A1
; GENERAL INFORMATION:
; APPLICANT: Warman, Matthew
; APPLICANT: Carpten, John
; APPLICANT: Trent, Jeffrey
; APPLICANT: Marcelino, Jose
; TITLE OF INVENTION: Novel Methods and Reagents for the Treatment of Osteoarthritis
; FILE REFERENCE: Case-06212
; CURRENT APPLICATION NUMBER: US/09/802,207
; CURRENT FILING DATE: 2001-08-29
; PRIOR APPLICATION NUMBER: 09/619,175
; PRIOR FILING DATE: 2000-07-19
; PRIOR APPLICATION NUMBER: 60/145,328
; PRIOR FILING DATE: 1999-07-23
; NUMBER OF SEQ ID NOS: 30
; SOFTWARE: Patent in version 3.0
; SEQ ID NO 30
; LENGTH: 1404
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-802-207-30

Query Match          99.8%; Score 5820.9; DB 9; Length 1404;
Best Local Similarity 96.4%; Pred. No. 6.6e-142;
Matches 1099; Conservative 0; Mismatches 0; Indels 41; Gaps 1;

QY 1 MAWKTLPIYLLLLSVFVIQVSSQ----- 25
DB 1 MAWKTLPIYLLLLSVFVIQVSSQDLSSCAGRGEGYSDATCNCYDNCQHYMECCPDF 60

QY 26 -----ELSCKGRCFESFERGECDDAQQCKYDKCCPDYEFCAEVHNPTSPSSSKAP 79
DB 61 KRVTAEELSCKGRCFESFERGECDDAQQCKYDKCCPDYEFCAEVHNPTSPSSSKAP 120

QY 80 PPSGASQTIKSTTKSPKPNKKTKVIESEIEEHSVSENQESSSSSSSSSTI 139
DB 121 PPSGASQTIKSTTKSPKPNKKTKVIESEIEEHSVSENQESSSSSSSSSTI 180

QY 140 KIKSSKNSAANRELQKLVKDKNKNRTKKKPTPKPPVVDGAGSLDNGDFKVTPTST 199
DB 181 KIKSSKNSAANRELQKLVKDKNKNRTKKKPTPKPPVVDGAGSLDNGDFKVTPTST 240

QY 200 TOHNKVSTSPKITTAKPINRPSLPNSDTSKETSLSLVNKKETVETKETTNNKQSTSDG 259
DB 241 TOHNKVSTSPKITTAKPINRPSLPNSDTSKETSLSLVNKKETVETKETTNNKQSTSDG 300

QY 260 KEKTTSAKETQSIKTSKADLAPTSKVLAKPTPKAETTTKGPALTTKPEPTTPKPEAS 319
DB 301 KEKTTSAKETQSIKTSKADLAPTSKVLAKPTPKAETTTKGPALTTKPEPTTPKPEAS 360

QY 320 TTPKEPTPTTIKSAPTTPKPEAPTTTKGAPTTKSPAPTTTKPEAPTTTKPEAPTTTKEP 379
DB 361 TTPKEPTPTTIKSAPTTPKPEAPTTTKGAPTTKSPAPTTTKPEAPTTTKPEAPTTTKEP 420

QY 380 APTTKSAPTTTKPEAPTTTKPEAPTTTKPEAPTTTKPEAPTTTKPEAPTTTKPEAPTPK 439
DB 421 APTTKSAPTTTKPEAPTTTKPEAPTTTKPEAPTTTKPEAPTTTKPEAPTTTKPEAPTPK 480

QY 440 EPAPTAPKKPAPTTTKPEAPTTTKPEAPTTTKPEAPTTTKPEAPTTTKSAPTTTKEPAPT 499
DB 481 EPAPTAPKKPAPTTTKPEAPTTTKPEAPTTTKPEAPTTTKPEAPTTTKSAPTTTKEPAPT 540

QY 500 TTKSAPTTTKPEAPTTTKPEAPTTTKPEAPTTTKPEAPTTTKPEAPTTTKPEAPTTTKKP 559
DB 541 TTKSAPTTTKPEAPTTTKPEAPTTTKPEAPTTTKPEAPTTTKPEAPTTTKPEAPTTTKKP 600

QY 560 APTAPKEAPTTTKPEAPTTTKPEAPTTTKPEAPTTTKPEAPTTTKPEAPTTTKPEAPTTT 619
DB 601 APTAPKEAPTTTKPEAPTTTKPEAPTTTKPEAPTTTKPEAPTTTKPEAPTTTKPEAPTTT 660

QY 620 PEEAPTTPKAAAPNTKPEAPTTTKPEAPTTTKPEAPTTTKPEAPTTTKPEAPTTTKPEAPTT 679
DB 661 PEEAPTTPKAAAPNTKPEAPTTTKPEAPTTTKPEAPTTTKPEAPTTTKPEAPTTTKPEAPTT 720

QY 680 APTTPKKAPKELAPTTTKPEPTSTSDKPAITTKGAPTTTKPEAPTTTKPEAPTTTKPG 739
DB 721 APTTPKKAPKELAPTTTKPEPTSTSDKPAITTKGAPTTTKPEAPTTTKPEAPTTTKPEAPTT 780

QY 740 TAPTTLKEPAPTTTKPEAPTTTKPEAPTTTKPEAPTTTKPEAPTTTKPEAPTTTKPEAPTT 799
DB 781 TAPTTLKEPAPTTTKPEAPTTTKPEAPTTTKPEAPTTTKPEAPTTTKPEAPTTTKPEAPTT 840

QY 800 KPAPTTPETPPPTTSVSTPTTKPTTIHKSPDESSTPELSAETPKALENSPKPGVPT 859
DB 841 KPAPTTPETPPPTTSVSTPTTKPTTIHKSPDESSTPELSAETPKALENSPKPGVPT 900

QY 860 TKTPAATKEMTTTAKDKTTERDLRTTPTTTTAAAPKMTKETATTTTEKTTESKITATTTQV 919
DB 901 TKTPAATKEMTTTAKDKTTERDLRTTPTTTTAAAPKMTKETATTTTEKTTESKITATTTQV 960

QY 920 TSTTTQDTPFPKITTLLKTTLLAPKVTTTKKTTTTTTEIMNKPBEETAKPKDRATNSKATTPK 979

```


RESULT 5
US-10-124-557-2
; Sequence 2, Application US/10124557
; Publication No. US20020137894A1
; GENERAL INFORMATION:
; APPLICANT: Turner, Katherine
; Clark, Stephen C.
; Jacobs, Kenneth
; Hewick, Rodney M.
; Gesner, Thomas G.
; TITLE OF INVENTION: Megakaryocyte Stimulating Factors
; NUMBER OF SEQUENCES: 143
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genetics Institute, Inc.
; STREET: 87 Cambridgepark Drive
; CITY: Cambridge
; STATE: Massachusetts
; COUNTRY: U.S.A.
; ZIP: 02140
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent in Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/10/124,557
; FILING DATE: 16-Apr-2002
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/643,502
; FILING DATE: 18-JAN-1991
; APPLICATION NUMBER: US 07/546,114
; FILING DATE: 29-JUN-1990
; APPLICATION NUMBER: US 07/457,196
; FILING DATE: 29-DEC-1989
; APPLICATION NUMBER: US 07/390,901
; FILING DATE: 08-AUG-1989
; ATTORNEY/AGENT INFORMATION:
; NAME: Cserr, Luann
; REGISTRATION NUMBER: 31,822
; REFERENCE/DOCKET NUMBER: GI 5190
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617) 876-1170
; TELEFAX: (617) 876-5851
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1404 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; SEQUENCE DESCRIPTION: SEQ ID NO: 2:
US-10-124-557-2
Query Match 99.8%; Score 5820.9; DB 13; Length 1404;
Best Local Similarity 96.4%; Pred. No. 6.6e-142;
Matches 1099; Conservative 0; Mismatches 0; Indels 41; Gaps 1;
QY 1 MAWKLPYLLLLLVFVIQVSSQ----- 25
DB 1 MAWKLPYLLLLLVFVIQVSSQDLSSACGRGEGYSRDATCNCYNCQHYMECCPDF 60
QY 26 -----ELSCGRGFESFERGECDCDAQCKYDKCCPDYEFCAEVHNPTSPSSKKAP 79
DB 61 KRVCTAEJSCGRGFESFERGECDCDAQCKYDKCCPDYEFCAEVHNPTSPSSKKAP 120
QY 80 PPSGASQTKSTKSPKPNKTKKVIIESEIIEHVSNSQSSSSSSSSSSSSSSSIW 139
DB 121 PPSGASQTKSTKSPKPNKTKKVIIESEIIEHVSNSQSSSSSSSSSSSSSIW 180
QY 140 KITSSKNSAANRELQKLVKDKNKKTKKTPKPPVVDAGSLDNGDFKVTTPDST 199

181 KIKSSKNSAANRELQKLVKDKNKKTKKTPKPPVVDAGSLDNGDFKVTTPDST 240
200 TOHNKVSTSPKITTAKPINRPSLPNSDTSKETSITVANKETTVETKETTTNKQSTDG 259
241 TOHNKVSTSPKITTAKPINRPSLPNSDTSKETSITVANKETTVETKETTTNKQSTDG 300
260 KEXTTSAKTOSIEKTSKDLAPTSKVLAKPTPKAETTTKGPALTTPKPPTTPPKEPAS 319
301 KEXTTSAKTOSIEKTSKDLAPTSKVLAKPTPKAETTTKGPALTTPKPPTTPPKEPAS 360
320 TTPKEPTPTTIKSAPTTTPKEPAPTTTKSAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTKP 379
361 TTPKEPTPTTIKSAPTTTPKEPAPTTTKSAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTKP 420
380 APTTTKSAPTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTKP 439
421 APTTTKSAPTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTKP 480
440 EPAPTAPKKPAPTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTKP 499
481 EPAPTAPKKPAPTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTKP 540
500 TTKSAPTTPKEPSPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTKP 559
541 TTKSAPTTPKEPSPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTKP 600
560 APTAPKEPAPTTPKETAPTTPKLTTPPKLAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTKP 619
601 APTAPKEPAPTTPKETAPTTPKLTTPPKLAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTKP 660
620 PEPAPPTPKAAAPNTPKPAPTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTKP 679
661 PEPAPPTPKAAAPNTPKPAPTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTKP 720
680 APTTPKKPAPKELAPTTPKEPTSTTSKAPTTTPKGTAPTTPKEPAPTTTPKEPAPTTTKP 739
721 APTTPKKPAPKELAPTTPKEPTSTTSKAPTTTPKGTAPTTPKEPAPTTTPKEPAPTTTKP 780
740 TAPTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTKP 799
781 TAPTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTKP 840
800 KPAPTTPPETPPTTSEVSTPTTTTKEPTTIHKSPDESTPELSAPPTPKALENSPKPQVPT 859
841 KPAPTTPPETPPTTSEVSTPTTTTKEPTTIHKSPDESTPELSAPPTPKALENSPKPQVPT 900
860 TKTPAATKPEVTTAKDKITERDLRTPETTTAAPKMTKETAATTTKTESKITATTQV 919
901 TKTPAATKPEVTTAKDKITERDLRTPETTTAAPKMTKETAATTTKTESKITATTQV 960
920 TSTTTQDTPPKITTLKTTTLAPKVTTKKTIITTEIMNKPEETAKPKDRATNSKATTPK 979
961 TSTTTQDTPPKITTLKTTTLAPKVTTKKTIITTEIMNKPEETAKPKDRATNSKATTPK 1020
980 POKPTKAPKXPTSTKPKMTMPRVKPKTTPTPRMTSTMPBLNPTSRIAEAMLOTTTRPN 1039
1021 POKPTKAPKXPTSTKPKMTMPRVKPKTTPTPRMTSTMPBLNPTSRIAEAMLOTTTRPN 1080
1040 QTNSKLVNPKSBDAGGAGETPHMLLRPHVFMPEVTPDMVLPVNPQGIINPMLS 1099
1081 QTNSKLVNPKSBDAGGAGETPHMLLRPHVFMPEVTPDMVLPVNPQGIINPMLS 1140

RESULT 6
US-10-124-557-62
; Sequence 62, Application US/10124557
; Publication No. US20020137894A1
; GENERAL INFORMATION:
; APPLICANT: Turner, Katherine
; Clark, Stephen C.
; Jacobs, Kenneth
; Hewick, Rodney M.
; Gesner, Thomas G.

TITLE OF INVENTION: Megakaryocyte Stimulating Factors
 NUMBER OF SEQUENCES: 143
 CORRESPONDENCE ADDRESS:

ADDRESSEE: Genetics Institute, Inc.
 STREET: 87 CambridgePark Drive
 CITY: Cambridge
 STATE: Massachusetts
 COUNTRY: U.S.A.
 ZIP: 02140

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: PatentIn Release #1.0, Version #1.25
 CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/10/124,557
 FILING DATE: 16-Apr-2002
 CLASSIFICATION: <Unknown>

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 07/643,502
 FILING DATE: 18-JAN-1991
 APPLICATION NUMBER: US 07/546,114
 FILING DATE: 29-JUN-1990
 APPLICATION NUMBER: US 07/457,196
 FILING DATE: 29-DEC-1989
 APPLICATION NUMBER: US 07/390,901
 FILING DATE: 08-AUG-1989

ATTORNEY/AGENT INFORMATION:

NAME: Cserr, Luann
 REGISTRATION NUMBER: 31,822
 REFERENCE/DOCKET NUMBER: GI 5190
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (617)876-1170
 TELEFAX: (617)876-5851

INFORMATION FOR SEQ ID NO: 62:

SEQUENCE CHARACTERISTICS:

LENGTH: 1404 amino acids

TYPE: amino acid

TOPOLOGY: linear

MOLECULE TYPE: protein

SEQUENCE DESCRIPTION: SEQ ID NO: 62:

US-10-124-557-62

Query Match

Best Local Similarity 99.8%; Score 5820.9; DB 13; Length 1404;

Matches 1099; Conservative 0; Mismatches 0; Indels 41; Gaps 1;

1 MAWKTLPIYLLLSVFIQVSSQ-----25

1 MAWKTLPIYLLLSVFIQVSSQDILSSCAGRCGYSRATCNCNDYNCQHYMECCPDF 60

26 -----ELSCGRGFESPERGECDCDAQCKYDKCCPDYESFCAEVHNPTSPSSKAP 79

61 KRVTAEELSCGRGFESPERGECDCDAQCKYDKCCPDYESFCAEVHNPTSPSSKAP 120

80 PPGASQTIKSTTKRSPKPNKKTIVIESEETEEHSVSENOESSSSSSSSSSSTI 139

121 PPGASQTIKSTTKRSPKPNKKTIVIESEETEEHSVSENOESSSSSSSSSSSTI 180

140 KIKSKNSAANRELQKLVKDNKNKTKKPTPKPPVDEAGSLDNGDFKVTTPDTST 199

181 KIKSKNSAANRELQKLVKDNKNKTKKPTPKPPVDEAGSLDNGDFKVTTPDTST 240

200 TQHNKVTSPKITTAKPINRPSLPKNSDTSKTSVKNKETTVEKTTTNKQSTDG 259

241 TQHNKVTSPKITTAKPINRPSLPKNSDTSKTSVKNKETTVEKTTTNKQSTDG 300

260 KEKTSKAKETOSIEKTSKADLAPTS KVLAKPTPKAETTTKGPALTTPKPEPAS 319

301 KEKTSKAKETOSIEKTSKADLAPTS KVLAKPTPKAETTTKGPALTTPKPEPAS 360

320 TTPKEPTPTTIKSAPTTPKPEAPTTTKSAPTTTKSAPTTTKSAPTTTKSAPTTTK 379

361 TTPKEPTPTTIKSAPTTPKPEAPTTTKSAPTTTKSAPTTTKSAPTTTKSAPTTTK 420
 380 APTTTKSAPTTPKPEAPTTTKKPAPTTPKPEAPTTTKPEPTPTTPKPEAPTTTKPE 439
 421 APTTTKSAPTTPKPEAPTTTKKPAPTTPKPEAPTTTKPEPTPTTPKPEAPTTTKPE 480
 440 EPAPTAPKPAPTTPKPEAPTTTKKPAPTTPKPEAPTTTKPEPTPTTPKPEAPTTTK 499
 481 EPAPTAPKPAPTTPKPEAPTTTKKPAPTTPKPEAPTTTKPEPTPTTPKPEAPTTTK 540
 500 TTKSAPTTPKPEPTPTTPKPEAPTTTKKPAPTTPKPEAPTTTKPEPTPTTPKPEAPTTTK 559
 541 TTKSAPTTPKPEPTPTTPKPEAPTTTKKPAPTTPKPEAPTTTKPEPTPTTPKPEAPTTTK 600
 560 APTAPKPEAPTTTKETAPTTPKLTPTTPEKLAAPTTPKPEAPTTTKPEPTPTTPKPE 619
 601 APTAPKPEAPTTTKETAPTTPKLTPTTPEKLAAPTTPKPEAPTTTKPEPTPTTPKPE 660
 620 PEEPAPTTPKAAAPNTPKPEAPTTTKPEAPTTTKPEAPTTTKPEPTPTTPKPEAPTTTK 679
 661 PEEPAPTTPKAAAPNTPKPEAPTTTKPEAPTTTKPEAPTTTKPEPTPTTPKPEAPTTTK 720
 680 APTTPKPAKPAKELAPTTTKPEPTSTTSDDKPAPTTPKGTAPTTPKPEAPTTTKPKG 739
 721 APTTPKPAKPAKELAPTTTKPEPTSTTSDDKPAPTTPKGTAPTTPKPEAPTTTKPKG 780
 740 TAPTTLKPEAPTTTPKPAKPAKELAPTTTKGTPTSTTSDDKPAPTTPKGTAPTTPKPEAPTTTK 799
 781 TAPTTLKPEAPTTTPKPAKPAKELAPTTTKGTPTSTTSDDKPAPTTPKGTAPTTPKPEAPTTTK 840
 800 KPAPTTPPEPTPTTPSEVSTPTTKPTTIHKSPDSESTPELSAEPPTPKALENSPKPEPVPT 859
 841 KPAPTTPPEPTPTTPSEVSTPTTKPTTIHKSPDSESTPELSAEPPTPKALENSPKPEPVPT 900
 860 TKTPAATKEPMTTAKDKTTREDLATTPTTTTAAPEKMTKETATTTTEKTESKITATTI 919
 901 TKTPAATKEPMTTAKDKTTREDLATTPTTTTAAPEKMTKETATTTTEKTESKITATTI 960
 920 TSTTTQDTPPFKITTLLKTTLLAPKVTITTKITITTEIMNKPEETAKPKDRATNSKATTPK 979
 961 TSTTTQDTPPFKITTLLKTTLLAPKVTITTKITITTEIMNKPEETAKPKDRATNSKATTPK 1020
 980 POKPTKAPKKTSTKPKTMRVRKPTTPPKMTSTMPBLNPTSRFAEAMLOTTIRPN 1039
 1021 POKPTKAPKKTSTKPKTMRVRKPTTPPKMTSTMPBLNPTSRFAEAMLOTTIRPN 1080
 1040 QTPNSKLVVNPKSDAGAGETPHMLLRPHVFMPEVTPMDYLPVRVNOGIINPMLS 1099
 1081 QTPNSKLVVNPKSDAGAGETPHMLLRPHVFMPEVTPMDYLPVRVNOGIINPMLS 1140

RESULT 7

US-10-124-557-46

; Sequence 46, Application US/10124557

; Publication No. US20020137894A1

; GENERAL INFORMATION:

; APPLICANT: Turner, Katherine

; Clark, Stephen C.

; Jacobs, Kenneth

; Hewick, Rodney M.

; Gesner, Thomas G.

; NUMBER OF SEQUENCES: 143

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Genetics Institute, Inc.

; STREET: 87 CambridgePark Drive

; CITY: Cambridge

; STATE: Massachusetts

; COUNTRY: U.S.A.

; ZIP: 02140

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible


```

; FILING DATE: 29-DEC-1989
; APPLICATION NUMBER: US 07/390,901
; FILING DATE: 08-AUG-1989
; ATTORNEY/AGENT INFORMATION:
;   NAME: Cserr, Luann
;   REGISTRATION NUMBER: 31,822
;   REFERENCE/DOCKET NUMBER: GI 5190
; TELECOMMUNICATION INFORMATION:
;   TELEPHONE: (617)876-1170
;   TELEFAX: (617)876-5851
; INFORMATION FOR SEQ ID NO: 60:
;   SEQUENCE CHARACTERISTICS:
;     LENGTH: 1320 amino acids
;     TYPE: amino acid
;     TOPOLOGY: linear
; MOLECULES TYPE: protein
; SEQUENCE DESCRIPTION: SEQ ID NO: 60:
US-10-124-557-60

Query Match
Best Local Similarity 96.3%; Score 5617.7; DB 13; Length 1320;
Matches 1056; Conservative 0; Mismatches 0; Indels 43; Gaps 1;

QY 1 MAWKTLPIYLLLVFVIQVSSQELSCKRCFESPERGECDCDAQCKKYDKCCPDYE 60
DB 1 MAWKTLPIYLLLVFVIQVSSQELSCKRCFESPERGECDCDAQCKKYDKCCPDYE 60
QY 61 SFCAVHNPTSPSSKAPPGSGASOTIKSTKSPKPNKKTKKVISEIEEHSVS 120
DB 61 SFCAVHNPTSPSSKAPPGSGASOTIKSTKSPKPNKKTKKVISEIEE----- 115
QY 121 ENQESSSSSSSSSTIWKIKSSKNSAANRELOKLVKDKNKKNRKTKKPTPPPVVDE 180
DB 116 -----VKDNKKNRKTKKPTPPPVVDE 137
QY 181 AGSGLDNGDFKVTTPDSTTOHNVKYSTSPKITTAKPINRPSLPNSDTSKETSITVKE 240
DB 138 AGSGLDNGDFKVTTPDSTTOHNVKYSTSPKITTAKPINRPSLPNSDTSKETSITVKE 197
QY 241 TVTETKETTITNKOTSDGKKTTSAXETQSIETSAKDLAPTSKVLAKPTPKAETTTKG 300
DB 198 TVTETKETTITNKOTSDGKKTTSAXETQSIETSAKDLAPTSKVLAKPTPKAETTTKG 257
QY 301 PALTPKEPTTPPKEPASTTPKEPTPTTIKSAPTTKPEAPTTKSAPTTPKPEAPTTT 360
DB 258 PALTPKEPTTPPKEPASTTPKEPTPTTIKSAPTTKPEAPTTKSAPTTPKPEAPTTT 317
QY 361 KEPAITPKPEAPTTKPEAPTTKSAPTTPKPEAPTTKPEAPTTKPEAPTTKPEPTP 420
DB 318 KEPAITPKPEAPTTKPEAPTTKSAPTTPKPEAPTTKPEAPTTKPEAPTTKPEPTP 377
QY 421 TTPKEAPTTKPEAPTTKPEAPTTAPKPAITPKPEAPTTKPEAPTTKPEAPTTTKE 480
DB 378 TTPKEAPTTKPEAPTTKPEAPTTAPKPAITPKPEAPTTKPEAPTTKPEAPTTTKE 437
QY 481 PAPTTIKSAPTTKPEAPTTKSAPTTPKESPTTKPEAPTTKPEAPTTKPEAPTTK 540
DB 438 PAPTTIKSAPTTKPEAPTTKSAPTTPKESPTTKPEAPTTKPEAPTTKPEAPTTK 497
QY 541 KEPAITPKPEAPTTTKKPAITAPKPAITPKPEAPTTPKKLTPTTPEKLAPTTPEKPA 600
DB 498 KEPAITPKPEAPTTTKKPAITAPKPAITPKPEAPTTPKKLTPTTPEKLAPTTPEKPA 557
QY 601 PTTPEELAPTTPEPTTPEEPAPTTPKAAAPNTPKEAPTTKPEAPTTKPEAPTTT 660
DB 558 PTTPEELAPTTPEPTTPEEPAPTTPKAAAPNTPKEAPTTKPEAPTTKPEAPTTT 617
QY 661 KETAPTTPKGTAPTTLKPEAPTTPKKAPKELAPTTTKPTSTTSKPAITPKGTAPTT 720
DB 618 KETAPTTPKGTAPTTLKPEAPTTPKKAPKELAPTTTKPTSTTSKPAITPKGTAPTT 677
QY 721 PKEPAPTTKPEAPTTKGTAPTTLKPEAPTTPKKAPKELAPTTTKGTSTTSKPAIT 780

```

RESULT 9

US-10-124-557-40

; Sequence 40, Application US/10124557

; Publication No. US2002037894A1

; GENERAL INFORMATION:

; APPLICANT: Turner, Katherine

; Clark, Stephen C.

; Jacobs, Kenneth

; Hewick, Rodney M.

; Gesner, Thomas G.

; TITLE OF INVENTION: Megakaryocyte Stimulating Factors

; NUMBER OF SEQUENCES: 143

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Genetix Institute, Inc.

; STREET: 87 Cambridgepark Drive

; CITY: Cambridge

; STATE: Massachusetts

; COUNTRY: U.S.A.

; ZIP: 02140

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: Patent in Release #1.0, Version #1.25

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/10/124,557

; FILING DATE: 16-Apr-2002

; CLASSIFICATION: <Unknown>

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: US 07/643,502

; FILING DATE: 18-JAN-1991

; APPLICATION NUMBER: US 07/546,114

; FILING DATE: 29-JUN-1990

; APPLICATION NUMBER: US 07/457,196

; FILING DATE: 29-DEC-1989

; APPLICATION NUMBER: US 07/390,901

; FILING DATE: 08-AUG-1989

; ATTORNEY/AGENT INFORMATION:

; NAME: Cserr, Luann

; REGISTRATION NUMBER: 31,822

; REFERENCE/DOCKET NUMBER: GI 5190

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: (617)876-1170

; TELEFAX: (617)876-5851

; INFORMATION FOR SEQ ID NO: 40:

; SEQUENCE CHARACTERISTICS:

```

;      LENGTH: 1361 amino acids
;      TYPE: amino acid
;      TOPOLOGY: linear
;      MOLECULE TYPE: protein
;      SEQUENCE DESCRIPTION: SEQ ID NO: 40:
US-10-124-557-40

Query Match          96.0%; Score 5603.6; DB 13; Length 1361;
Best Local Similarity 92.6%; Pred. No. 2.4e-136;
Matches 1056; Conservative 0; Mismatches 0; Indels 84; Gaps 2;

QY 1 MAWKTLPIYLLLLSVFVIQVSSQ----- 25
DB 1 MAWKTLPIYLLLLSVFVIQVSSQDLSSCAGRCGEGYRSDATCNDYNCQHMECCPDF 60

QY 26 -----ELSKGRCFESFERGECDCDAQCKYDKCCDPDYRSFCAEVNPTSPSSKKAP 79
DB 61 KSVTAELSKGRCFESFERGECDCDAQCKYDKCCDPDYRSFCAEVNPTSPSSKKAP 120

QY 80 PPSGASQIKSTTKSPKPNKKTKVIESEITEEHSVSENQSSSSSSSSSSSSSIW 139
DB 121 PPSGASQIKSTTKSPKPNKKTKVIESEITE----- 156

QY 140 KIKSSKNSAANRELQKKLVKDNKNRKKKTPPKPPVVDVDEAGSLDNGDFKVTPTDST 199
DB 157 -----VKDNKNRKKKTPPKPPVVDVDEAGSLDNGDFKVTPTDST 197

QY 200 TQHNKVSPTKITTAKPINRPSLPNSDTSKETSITVNKETTVETKETTINKOTSDG 259
DB 198 TQHNKVSPTKITTAKPINRPSLPNSDTSKETSITVNKETTVETKETTINKOTSDG 257

QY 260 KEKTTSAKETQSIKTSADLAPTSKVLAKPTPKAETTTKGPAITTPKEPTTPPKEPAS 319
DB 258 KEKTTSAKETQSIKTSADLAPTSKVLAKPTPKAETTTKGPAITTPKEPTTPPKEPAS 317

QY 320 TTPKEPTPTIISAPTTKEPAPTTTKSAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEP 379
DB 318 TTPKEPTPTIISAPTTKEPAPTTTKSAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEP 377

QY 380 APTTTKSAPTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTT 439
DB 378 APTTTKSAPTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTT 437

QY 440 EPAPTKAPKAPTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKSAPTTTKEPAPT 499
DB 438 EPAPTKAPKAPTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKSAPTTTKEPAPT 497

QY 500 TTKSAPTTKEPSPTTKEPAPTTTKEPAPTTTKAPATTPKEPAPTTTKEPAPTTTKKP 559
DB 498 TTKSAPTTKEPSPTTKEPAPTTTKEPAPTTTKAPATTPKEPAPTTTKEPAPTTTKKP 557

QY 560 APTAPKEPAPTTKETAPTTPKLPTTPEKLAPTTPEKLAPTTPEELAPTTPPEPTPTT 619
DB 558 APTAPKEPAPTTKETAPTTPKLPTTPEKLAPTTPEKLAPTTPEELAPTTPPEPTPTT 617

QY 620 PREPAPTTKAAAPNTPEPAPTTKEPAPTTKEPAPTTKEPAPTTKETAPTTPKGAPATLKEP 679
DB 618 PREPAPTTKAAAPNTPEPAPTTKEPAPTTKEPAPTTKEPAPTTKETAPTTPKGAPATLKEP 677

QY 680 APTTPKAPKELAPTTTKEPTSTSDKAPTTTKGAPTTTKGAPTTTKGAPTTTKGAPTTPKG 739
DB 678 APTTPKAPKELAPTTTKEPTSTSDKAPTTTKGAPTTTKGAPTTTKGAPTTTKGAPTTPKG 737

QY 740 TAPTTLKEPAPTTPKAPKELAPTTTKGPTSTSDKAPTTTKGAPTTTKGAPTTTKGAPTTPK 799
DB 738 TAPTTLKEPAPTTPKAPKELAPTTTKGPTSTSDKAPTTTKGAPTTTKGAPTTTKGAPTTPK 797

QY 800 KPAPTTPEPTPTTSVSTPTTKEPTTIHKSPDSTPELSAEPPTKALENSPKPGVPT 859
DB 798 KPAPTTPEPTPTTSVSTPTTKEPTTIHKSPDSTPELSAEPPTKALENSPKPGVPT 857

QY 860 TKTPAAKPEMTTAKDKTTERDLRTTPTTTAAAPKMTKETATTTTEKTESKITATTQV 917
DB 858 TKTPAAKPEMTTAKDKTTERDLRTTPTTTAAAPKMTKETATTTTEKTESKITATTQV 917

QY 920 TSTTTQDTPPKITTLTKITTLAPKVTITTKITITTEIMKPEETAKPKORATNSKATTPK 979
DB 918 TSTTTQDTPPKITTLTKITTLAPKVTITTKITITTEIMKPEETAKPKORATNSKATTPK 977

QY 980 PQKPTKAPKKPTSTKKPKMTPRVKPKTTPTRKMTSTMPLNPTSTRIAEAMLQTTTRPN 1039
DB 978 PQKPTKAPKKPTSTKKPKMTPRVKPKTTPTRKMTSTMPLNPTSTRIAEAMLQTTTRPN 1037

QY 1040 OTNSKLVNPNKSEDAAGAGETPHMLLRPHVFMDEVTPDMVDYLRVFNQGIINPMLS 1099
DB 1038 OTNSKLVNPNKSEDAAGAGETPHMLLRPHVFMDEVTPDMVDYLRVFNQGIINPMLS 1097

RESULT 10
US-10-124-557-58
; Sequence 58, Application US/10124557
; Publication No. US20020137894A1
; GENERAL INFORMATION:
; APPLICANT: Turner, Katherine
;               Clark, Stephen C.
;               Jacobs, Kenneth M.
;               Hewick, Rodney M.
;               Gesner, Thomas G.
; TITLE OF INVENTION: Megakaryocyte Stimulating Factors
; NUMBER OF SEQUENCES: 143
; CORRESPONDENCE ADDRESSES:
; ADDRESSEE: Genetics Institute, Inc.
; STREET: 87 CambridgePark Drive
; CITY: Cambridge
; STATE: Massachusetts
; COUNTRY: U.S.A.
; ZIP: 02140
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/10/124,557
; FILING DATE: 16-Apr-2002
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/643,502
; FILING DATE: 18-JAN-1991
; APPLICATION NUMBER: US 07/546,114
; FILING DATE: 29-JUN-1990
; APPLICATION NUMBER: US 07/457,196
; FILING DATE: 29-DEC-1989
; APPLICATION NUMBER: US 07/390,901
; FILING DATE: 08-AUG-1989
; ATTORNEY/AGENT INFORMATION:
; NAME: Cserr, Luann
; REGISTRATION NUMBER: 31,822
; REFERENCE/DOCKET NUMBER: GI 5190
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617)876-1170
; TELEFAX: (617)876-5851
; INFORMATION FOR SEQ ID NO: 58:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1049 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; SEQUENCE DESCRIPTION: SEQ ID NO: 58:
US-10-124-557-58

Query Match          95.3%; Score 5561; DB 13; Length 1049;
Best Local Similarity 95.5%; Pred. No. 2.3e-135;
Matches 1049; Conservative 0; Mismatches 0; Indels 50; Gaps 1;

QY 1 MAWKTLPIYLLLLSVFVIQVSSQELSCGRCFESFERGECDCDAQCKYDKCCPDFE 60
```

```
Db 1 MAWKTLPIYLLLLSVFVIQVSSQELSCKGRCFESFERGECDCDAQCKYDKCCPDYE 60
QY 61 SCAEVHNTSPSSKKAPPPGASQTIKSTTKRSPKPNKKTKKVIIESEITEHSVS 120
Db 61 SCA-----EHSVS 70
QY 121 ENQESSSSSSSSSTIWKIKSSKNSAANRELQKLVKDKNKRKTKKTPKPPVVD 180
Db 71 ENQESSSSSSSSSTIWKIKSSKNSAANRELQKLVKDKNKRKTKKTPKPPVVD 130
QY 181 AGSLDNGDFKVTTPDTSTQHNKYSTPKITAKPINRPSLPNSDTSKETSILTVNKE 240
Db 131 AGSLDNGDFKVTTPDTSTQHNKYSTPKITAKPINRPSLPNSDTSKETSILTVNKE 190
QY 241 TTIVETKETTNNKQSTDKGKETSASKEQTSIEKTSADLAPTSKVLAKPTPKAETTTKG 300
Db 191 TTIVETKETTNNKQSTDKGKETSASKEQTSIEKTSADLAPTSKVLAKPTPKAETTTKG 250
QY 301 PALTTPKETPTTPKEPASTTPKEPTPTTIKSAPTTPKEPAPTTPKSPAPTTT 360
Db 251 PALTTPKETPTTPKEPASTTPKEPTPTTIKSAPTTPKEPAPTTPKSPAPTTT 310
QY 361 KEPAATTPKEPAPTTPKEPAPTTPKSAPTTPKEPAPTTPKSPAPTTTPKEPTT 420
Db 311 KEPAATTPKEPAPTTPKEPAPTTPKSAPTTPKEPAPTTPKSPAPTTTPKEPTT 370
QY 421 TTPKEPAPTTPKEPAPTTPKEPAPTTPKSPAPTTTPKSPAPTTTPKEPTTPKE 480
Db 371 TTPKEPAPTTPKEPAPTTPKEPAPTTPKSAPTTPKEPAPTTPKSPAPTTTPKE 430
QY 481 PAPTTKSAPTTPKEPAPTTPKSAPTTPKSPAPTTTPKSPAPTTTPKEPAPTTP 540
Db 431 PAPTTKSAPTTPKEPAPTTPKSAPTTPKSPAPTTTPKSPAPTTTPKEPAPTTP 490
QY 541 KEPAATTPKEPAPTTPKAPAPAPAPAPAPAPAPAPAPAPAPAPAPAPAPAPAP 600
Db 491 KEPAATTPKEPAPTTPKAPAPAPAPAPAPAPAPAPAPAPAPAPAPAPAPAPAP 550
QY 601 PTPBELAPTTPEPTPTTPKEPAPTTPKAAAPNTPKEPAPTTPKSPAPTTTPKE 660
Db 551 PTPBELAPTTPEPTPTTPKEPAPTTPKAAAPNTPKEPAPTTPKSPAPTTTPKE 610
QY 661 KETAPTPKGTAPTTKEPAPTTPKAPAPAPAPAPAPAPAPAPAPAPAPAPAP 720
Db 611 KETAPTPKGTAPTTKEPAPTTPKAPAPAPAPAPAPAPAPAPAPAPAPAPAP 670
QY 721 PKEPAPTTPKEPAPTTPKGTAPTTKEPAPTTPKAPAPAPAPAPAPAPAPAPAP 780
Db 671 PKEPAPTTPKEPAPTTPKGTAPTTKEPAPTTPKAPAPAPAPAPAPAPAPAPAP 730
QY 781 TPKEPAPTTPKEPAPTTPKAPAPAPAPAPAPAPAPAPAPAPAPAPAPAPAPAP 840
Db 731 TPKEPAPTTPKEPAPTTPKAPAPAPAPAPAPAPAPAPAPAPAPAPAPAPAPAP 790
QY 841 AEPTKALENSPKPGVPTTTPKAPAPAPAPAPAPAPAPAPAPAPAPAPAPAPAP 900
Db 791 AEPTKALENSPKPGVPTTTPKAPAPAPAPAPAPAPAPAPAPAPAPAPAPAPAP 850
QY 901 ATTTEKTTESKITATTQVTSITQDITTPPKITTLKTLTILAPKVTTKITTEIMNKP 960
Db 851 ATTTEKTTESKITATTQVTSITQDITTPPKITTLKTLTILAPKVTTKITTEIMNKP 910
QY 961 BETAKPKDRATNSKATTPKOKPKAPKPKPTSTKPKTMPVRVKPTTPPKMTSTWPE 1020
Db 911 BETAKPKDRATNSKATTPKOKPKAPKPKPTSTKPKTMPVRVKPTTPPKMTSTWPE 970
QY 1021 LNPTSRIAEAMLOTTTRNQNTFNKLVNPKSDAGABGETPHMLLRPHVFMPEVTPD 1080
Db 971 LNPTSRIAEAMLOTTTRNQNTFNKLVNPKSDAGABGETPHMLLRPHVFMPEVTPD 1030
QY 1081 MDYLPRVNPQIINPMLS 1099
Db 1099 MDYLPRVNPQIINPMLS 1049
```

```
Db 1031 MDYLPRVNPQIINPMLS 1049
RESULT 11
US-10-124-557-142
; Sequence 142, Application US/10124557
; Publication No. US20020137894A1
; GENERAL INFORMATION:
; APPLICANT: Turner, Katherine
; Clark, Stephen C.
; Jacobs, Kenneth
; Hewick, Rodney M.
; Gesner, Thomas G.
; TITLE OF INVENTION: Megakaryocyte Stimulating Factors
; NUMBER OF SEQUENCES: 143
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genetics Institute, Inc.
; STREET: 87 CambridgePark Drive
; CITY: Cambridge
; STATE: Massachusetts
; COUNTRY: U.S.A.
; ZIP: 02140
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/10/124,557
; FILING DATE: 16-Apr-2002
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/643,502
; FILING DATE: 18-JAN-1991
; APPLICATION NUMBER: US 07/546,114
; FILING DATE: 29-JUN-1990
; APPLICATION NUMBER: US 07/457,196
; FILING DATE: 29-DEC-1989
; APPLICATION NUMBER: US 07/390,901
; FILING DATE: 08-AUG-1989
; ATTORNEY/AGENT INFORMATION:
; NAME: Cseri, Luann
; REGISTRATION NUMBER: 31,822
; REFERENCE/DOCKET NUMBER: GI 5190
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617)876-1170
; TELEFAX: (617)876-5851
; INFORMATION FOR SEQ ID NO: 142:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1313 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; SEQUENCE DESCRIPTION: SEQ ID NO: 142:
US-10-124-557-142
Query Match 95.3%; Score 5561; DB 13; Length 1313;
Best Local Similarity 95.5%; Pred. No. 2.9e-135;
Matches 1049; Conservative 0; Mismatches 0; Indels 50; Gaps 1;
QY 1 MAWKTLPIYLLLLSVFVIQVSSQELSCKGRCFESFERGECDCDAQCKYDKCCPDYE 60
Db 1 MAWKTLPIYLLLLSVFVIQVSSQELSCKGRCFESFERGECDCDAQCKYDKCCPDYE 60
QY 61 SCAEVHNTSPSSKKAPPPGASQTIKSTTKRSPKPNKKTKKVIIESEITEHSVS 120
Db 61 SCA-----EHSVS 70
QY 121 ENQESSSSSSSSSTIWKIKSSKNSAANRELQKLVKDKNKRKTKKTPKPPVVD 180
Db 71 ENQESSSSSSSSSTIWKIKSSKNSAANRELQKLVKDKNKRKTKKTPKPPVVD 130
QY 181 AGSLDNGDFKVTTPDTSTQHNKYSTPKITAKPINRPSLPNSDTSKETSILTVNKE 240
Db 131 AGSLDNGDFKVTTPDTSTQHNKYSTPKITAKPINRPSLPNSDTSKETSILTVNKE 190
QY 241 TTIVETKETTNNKQSTDKGKETSASKEQTSIEKTSADLAPTSKVLAKPTPKAETTTKG 300
Db 191 TTIVETKETTNNKQSTDKGKETSASKEQTSIEKTSADLAPTSKVLAKPTPKAETTTKG 250
QY 301 PALTTPKETPTTPKEPASTTPKEPTPTTIKSAPTTPKEPAPTTPKSPAPTTT 360
Db 251 PALTTPKETPTTPKEPASTTPKEPTPTTIKSAPTTPKEPAPTTPKSPAPTTT 310
QY 361 KEPAATTPKEPAPTTPKEPAPTTPKSAPTTPKEPAPTTPKSPAPTTTPKEPTT 420
Db 311 KEPAATTPKEPAPTTPKEPAPTTPKSAPTTPKEPAPTTPKSPAPTTTPKEPTT 370
QY 421 TTPKEPAPTTPKEPAPTTPKEPAPTTPKSPAPTTTPKSPAPTTTPKEPTTPKE 480
Db 371 TTPKEPAPTTPKEPAPTTPKEPAPTTPKSAPTTPKEPAPTTPKSPAPTTTPKE 430
QY 481 PAPTTKSAPTTPKEPAPTTPKSAPTTPKSPAPTTTPKSPAPTTTPKEPAPTTP 540
Db 431 PAPTTKSAPTTPKEPAPTTPKSAPTTPKSPAPTTTPKSPAPTTTPKEPAPTTP 490
QY 541 KEPAATTPKEPAPTTPKAPAPAPAPAPAPAPAPAPAPAPAPAPAPAPAPAPAP 600
Db 491 KEPAATTPKEPAPTTPKAPAPAPAPAPAPAPAPAPAPAPAPAPAPAPAPAPAP 550
QY 601 PTPBELAPTTPEPTPTTPKEPAPTTPKAAAPNTPKEPAPTTPKSPAPTTTPKE 660
Db 551 PTPBELAPTTPEPTPTTPKEPAPTTPKAAAPNTPKEPAPTTPKSPAPTTTPKE 610
QY 661 KETAPTPKGTAPTTKEPAPTTPKAPAPAPAPAPAPAPAPAPAPAPAPAPAP 720
Db 611 KETAPTPKGTAPTTKEPAPTTPKAPAPAPAPAPAPAPAPAPAPAPAPAPAP 670
QY 721 PKEPAPTTPKEPAPTTPKGTAPTTKEPAPTTPKAPAPAPAPAPAPAPAPAPAP 780
Db 671 PKEPAPTTPKEPAPTTPKGTAPTTKEPAPTTPKAPAPAPAPAPAPAPAPAPAP 730
QY 781 TPKEPAPTTPKEPAPTTPKAPAPAPAPAPAPAPAPAPAPAPAPAPAPAPAPAP 840
Db 731 TPKEPAPTTPKEPAPTTPKAPAPAPAPAPAPAPAPAPAPAPAPAPAPAPAPAP 790
QY 841 AEPTKALENSPKPGVPTTTPKAPAPAPAPAPAPAPAPAPAPAPAPAPAPAPAP 900
Db 791 AEPTKALENSPKPGVPTTTPKAPAPAPAPAPAPAPAPAPAPAPAPAPAPAPAP 850
QY 901 ATTTEKTTESKITATTQVTSITQDITTPPKITTLKTLTILAPKVTTKITTEIMNKP 960
Db 851 ATTTEKTTESKITATTQVTSITQDITTPPKITTLKTLTILAPKVTTKITTEIMNKP 910
QY 961 BETAKPKDRATNSKATTPKOKPKAPKPKPTSTKPKTMPVRVKPTTPPKMTSTWPE 1020
Db 911 BETAKPKDRATNSKATTPKOKPKAPKPKPTSTKPKTMPVRVKPTTPPKMTSTWPE 970
QY 1021 LNPTSRIAEAMLOTTTRNQNTFNKLVNPKSDAGABGETPHMLLRPHVFMPEVTPD 1080
Db 971 LNPTSRIAEAMLOTTTRNQNTFNKLVNPKSDAGABGETPHMLLRPHVFMPEVTPD 1030
QY 1081 MDYLPRVNPQIINPMLS 1099
Db 1099 MDYLPRVNPQIINPMLS 1049
```

Db 131 AGSLDNGDFKVTPTDSTQHNKVSSTSPKITTAKPINRPSLPNSDTSKETSITVAK 190
Qy 241 TTVETKETTTHKQSTGKKTSAKETSQSTKSAKDLAPTSKVLAKPTPKATTTKG 300
Db 191 TTVETKETTTHKQSTGKKTSAKETSQSTKSAKDLAPTSKVLAKPTPKATTTKG 250
Qy 301 PALTTPKETPTTTPKEPASTTTPKEPTTTIKSAPTTPKEPAPTTTTSKAPTTPKEPAPTTT 360
Db 251 PALTTPKETPTTTPKEPASTTTPKEPTTTIKSAPTTPKEPAPTTTTSKAPTTPKEPAPTTT 310
Qy 361 KEPAATTPKEPAPTTTTPKEPASTTTPKEPTTTIKSAPTTPKEPAPTTTTSKAPTTPKEPAPTTT 420
Db 311 KEPAATTPKEPAPTTTTPKEPASTTTPKEPTTTIKSAPTTPKEPAPTTTTSKAPTTPKEPAPTTT 370
Qy 421 TTPKEPAPTTTPKEPAPTTTPKEPASTTTPKEPTTTIKSAPTTPKEPAPTTTTSKAPTTPKEPAPTTT 480
Db 371 TTPKEPAPTTTPKEPAPTTTPKEPASTTTPKEPTTTIKSAPTTPKEPAPTTTTSKAPTTPKEPAPTTT 430
Qy 481 PAPTTPKSAPTTTPKEPAPTTTTPKEPASTTTPKEPTTTIKSAPTTPKEPAPTTTTSKAPTTPKEPAPTTT 540
Db 431 PAPTTPKSAPTTTPKEPAPTTTTPKEPASTTTPKEPTTTIKSAPTTPKEPAPTTTTSKAPTTPKEPAPTTT 490
Qy 541 KEPAATTPKEPAPTTTTPKEPASTTTPKEPTTTIKSAPTTPKEPAPTTTTSKAPTTPKEPAPTTTTPKEPA 600
Db 491 KEPAATTPKEPAPTTTTPKEPASTTTPKEPTTTIKSAPTTPKEPAPTTTTSKAPTTPKEPAPTTTTPKEPA 550
Qy 601 PTPPELAPTTPEPTTPKEPAPTTTPKEPASTTTPKEPTTTIKSAPTTPKEPAPTTTTSKAPTTPKEPAPTTT 660
Db 551 PTPPELAPTTPEPTTPKEPAPTTTPKEPASTTTPKEPTTTIKSAPTTPKEPAPTTTTSKAPTTPKEPAPTTT 610
Qy 661 KETAPTTPKGTAPTTPKEPAPTTTPKEPASTTTPKEPTTTIKSAPTTPKEPAPTTTTSKAPTTPKEPAPTTT 720
Db 611 KETAPTTPKGTAPTTPKEPAPTTTPKEPASTTTPKEPTTTIKSAPTTPKEPAPTTTTSKAPTTPKEPAPTTT 670
Qy 721 PKEPAPTTTPKEPAPTTTPKEPASTTTPKEPTTTIKSAPTTPKEPAPTTTTSKAPTTPKEPAPTTTTPKEPAP 780
Db 671 PKEPAPTTTPKEPAPTTTPKEPASTTTPKEPTTTIKSAPTTPKEPAPTTTTSKAPTTPKEPAPTTTTPKEPAP 730
Qy 781 TPKEAPTTPKEPAPTTTPKEPASTTTPKEPTTTIKSAPTTPKEPAPTTTTSKAPTTPKEPAPTTTTPKEPAP 840
Db 731 TPKEAPTTPKEPAPTTTPKEPASTTTPKEPTTTIKSAPTTPKEPAPTTTTSKAPTTPKEPAPTTTTPKEPAP 790
Qy 841 ABPTPKALENSPKPGVPTTKTAAKTPKEMTTTAKDKTERDLRTTTPETTTAAPKMTKET 900
Db 791 ABPTPKALENSPKPGVPTTKTAAKTPKEMTTTAKDKTERDLRTTTPETTTAAPKMTKET 850
Qy 901 ATTEKTTESKITATTTQVSTTTQDTPFKITTLKTTTLAPKVTTTKKTTITTEIMNKP 960
Db 851 ATTEKTTESKITATTTQVSTTTQDTPFKITTLKTTTLAPKVTTTKKTTITTEIMNKP 910
Qy 961 ETAKPKDRAKTSKATTPKQKPTKAPKPTSTKPKMTVRKPKTTPTPKMTSTWPE 1020
Db 911 ETAKPKDRAKTSKATTPKQKPTKAPKPTSTKPKMTVRKPKTTPTPKMTSTWPE 970
Qy 1021 LNPTSRIBAMLOTTTRPNQTPNSKLVEVNPXSEDAGAGETPHMLRPHVFMPEVTPD 1080
Db 971 LNPTSRIBAMLOTTTRPNQTPNSKLVEVNPXSEDAGAGETPHMLRPHVFMPEVTPD 1030
Qy 1081 MDYLPVNPQGIINPMLS 1099
Db 1031 MDYLPVNPQGIINPMLS 1049

RESULT 12
US-10-124-557-48
; Sequence 48, Application US/10124557
; Publication No. US20020137894A1
; GENERAL INFORMATION:
; APPLICANT: Turner, Katherine
; Clark, Stephen C.
; Jacobs, Kenneth
; Hewick, Rodney M.

Gesner, Thomas G.
TITLE OF INVENTION: Megakaryocyte Stimulating Factors
NUMBER OF SEQUENCES: 143
CORRESPONDENCE ADDRESS:
ADDRESSEE: Genetics Institute, Inc.
STREET: 87 Cambridgepark Drive
CITY: Cambridge
STATE: Massachusetts
COUNTRY: U.S.A.
ZIP: 02140
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/10/124,557
FILING DATE: 16-APR-2002
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/643,502
FILING DATE: 18-JAN-1991
APPLICATION NUMBER: US 07/546,114
FILING DATE: 29-JUN-1990
APPLICATION NUMBER: US 07/457,196
FILING DATE: 29-DEC-1989
APPLICATION NUMBER: US 07/390,901
FILING DATE: 08-AUG-1989
ATTORNEY/AGENT INFORMATION:
NAME: Cseerr, Luann
REGISTRATION NUMBER: 31,822
REFERENCE/DOCKET NUMBER: GI 5190
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617)876-1170
TELEFAX: (617)876-5851
INFORMATION FOR SEQ ID NO: 48:
SEQUENCE CHARACTERISTICS:
LENGTH: 1354 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
SEQUENCE DESCRIPTION: SEQ ID NO: 48:
US-10-124-557-48

Query Match 95.1%; Score 5546.9; DB 13; Length 1354;
Best Local Similarity 92.0%; Pred. No. 7e-135;
Matches 1049; Conservative 0; Mismatches 0; Indels 91; Gaps 2;
Qy 1 MAWKTLPIYLLLSLVFVIQVSSQ----- 25
Db 1 MAWKTLPIYLLLSLVFVIQVSSQDLS SCAGRCGYSRDATCNCDCYNQHYMECCPDF 60
Qy 26 -----ELSCKGRCFESFEGRECDCAQCKYDKCPDYESFCAEYHNFTSPSSKKAP 79
Db 61 KRVCTAELSCKGRCFESFEGRECDCAQCKYDKCPDYESFCA----- 105
Qy 80 PPASGASQTIKSTTKRSPKPPNKKTKKVIKIESEBITEHSVSENEQSSSSSSSSSSSTIW 139
Db 106 -----EEHSVSENEQSSSSSSSSSSSSSTIW 130
Qy 140 KIKSKNSAANRELOKXKLVKDNKNKNTKKKPTKPPVVDVDEAGSLDNGDFKVTPTDST 199
Db 131 KIKSKNSAANRELOKXKLVKDNKNKNTKKKPTKPPVVDVDEAGSLDNGDFKVTPTDST 190
Qy 200 TQHNKVSPTSKITTAKEPINRPSLPNSDTSKETSITVAKETTVEKETTITNKQTSIDG 259
Db 191 TQHNKVSPTSKITTAKEPINRPSLPNSDTSKETSITVAKETTVEKETTITNKQTSIDG 250
Qy 260 KEKTTSAKETQSIEKTSKDLAPTSKVLAKPTPKAETTTKGPALTTTPKEPTTPKEPAS 319
Db 251 KEKTTSAKETQSIEKTSKDLAPTSKVLAKPTPKAETTTKGPALTTTPKEPTTPKEPAS 310
Qy 320 TTPKEPTTPKSAPTTPKEPAPTTTTPKEPAPTTTTPKEPAPTTTTPKEPAPTTTTPKEP 379

| | | | |
|------|----|--|------|
| 311 | Db | TTTKEPTPTTIIKSAPTTTPKBPAPTTTTSABTTTPKEPAPTTTKEPAPTTTPKEPAPTTTKEP | 370 |
| 380 | Qy | APTITTKSAPTTTPKEPAPTTTPKPKAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKBPAPTTTK | 439 |
| 371 | Db | APTITTKSAPTTTPKEPAPTTTPKPKAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTK | 430 |
| 440 | Qy | EPAPTAPKBPAPTTTPKEPAPTTTPKEPAPTTTTPKPSPTTPKEPAPTTTTPKSAPTTTKEPAPT | 499 |
| 431 | Db | EPAPTAPKBPAPTTTPKEPAPTTTPKEPAPTTTTPKPSPTTPKEPAPTTTTPKSAPTTTKEPAPT | 490 |
| 500 | Qy | TTTSAPTTTPKEPSPTTTPKBPAPTTTPKBPAPTTTPKPKAPTTTPKBPAPTTTPKEPAPTTTPKPK | 559 |
| 491 | Db | TTTSAPTTTPKEPSPTTTPKBPAPTTTPKBPAPTTTPKPKAPTTTPKBPAPTTTPKEPAPTTTPKPK | 550 |
| 560 | Qy | APTAPKEPAPTTTPKETAPTTTPKKLTPTTTPBKLAPTTPKEKAPTTTPPELAPTTTPPEPTPTT | 619 |
| 551 | Db | APTAPKEPAPTTTPKETAPTTTPKKLTPTTTPBKLAPTTPKEKAPTTTPPELAPTTTPPEPTPTT | 610 |
| 620 | Qy | PEBPAPTTTPKAAAPNTTPKEPAPTTTPKBPAPTTTPKEPAPTTTPKETAPTTTPKGTAPTTIKEP | 679 |
| 611 | Db | PEBPAPTTTPKAAAPNTTPKEPAPTTTPKBPAPTTTPKEPAPTTTPKETAPTTTPKGTAPTTIKEP | 670 |
| 680 | Qy | APTTPKBPAPKELAPTTTTPKSTSTTSOKPAPTTTPKGTAPTTTPKEPAPTTTPKEPAPTTTPKPG | 739 |
| 671 | Db | APTTPKBPAPKELAPTTTTPKSTSTTSOKPAPTTTPKGTAPTTTPKEPAPTTTPKEPAPTTTPKPG | 730 |
| 740 | Qy | TAPTTLKEPAPTTTPKBPAPKELAPTTTTPKGTSTTSOKPAPTTTPKETAPTTTPKBPAPTTTPK | 799 |
| 731 | Db | TAPTTLKEPAPTTTPKBPAPKELAPTTTTPKGTSTTSOKPAPTTTPKETAPTTTPKBPAPTTTPK | 790 |
| 800 | Qy | KPAPTTPETPPPTTSEVSTPTTTPKEPTTIHKSPDESTPELSAPTPPKALSNSPKPCVPT | 859 |
| 791 | Db | KPAPTTPETPPPTTSEVSTPTTTPKEPTTIHKSPDESTPELSAPTPPKALSNSPKPCVPT | 850 |
| 860 | Qy | TKTPAAATPKBMTTAKDKTTTERDLRTPTTPTTAAPKMTKETATTTTKEKTESKITATTTQV | 919 |
| 851 | Db | TKTPAAATPKBMTTAKDKTTTERDLRTPTTPTTAAPKMTKETATTTTKEKTESKITATTTQV | 910 |
| 920 | Qy | TSITTQDTPPTFKITTLKITTLAPKVTTTTKXITITTEBIMNKPEETAKPKDRATNSKATTPK | 979 |
| 911 | Db | TSITTQDTPPTFKITTLKITTLAPKVTTTTKXITITTEBIMNKPEETAKPKDRATNSKATTPK | 970 |
| 980 | Qy | PQKPTKAPKPTSTPKPKTWPVRVKPXTTTPPKQWSTTMPPELNPTTSIAEAMLOTTTRPN | 1039 |
| 971 | Db | PQKPTKAPKPTSTPKPKTWPVRVKPXTTTPPKQWSTTMPPELNPTTSIAEAMLOTTTRPN | 1030 |
| 1040 | Qy | QTPNSKLIVEVNPXSIEDAGGAETPHMLLSRPHVMPVPTTDMDYLPRAVPVNOGIIINPMLS | 1099 |
| 1031 | Db | QTPNSKLIVEVNPXSIEDAGGAETPHMLLSRPHVMPVPTTDMDYLPRAVPVNOGIIINPMLS | 1090 |

RESULT 13
US-10-124-557-50

US-10-124-557-50
; Sequence 50, Application US/10124557
: Publication No. US20020137894A1

Publication No. US20020000000

APPLICANT: Turner, Katherine
Clark, Stephen C.
Jacobs, Kenneth
Hewick, Rodney M.
Gesner, Thomas G.

Geisner, Thomas G.
TITLE OF INVENTION: Megakaryocyte Stimulating Factors

TITLE OF INVENTION: Mega
 NUMBER OF SEQUENCES: 143

CORRESPONDENCE ADDRESS:

ADDRESSEE: Genetics Institute, Inc.

STREET: 87 Cambridgepark Drive

STREET: 87 Cambridge
CITY: Cambridge

CITY: Cambridge
STATE: Massachusetts

COUNTRY: U.S.A.

COUNTY: 0
ZIP: 02140

DATE: 02140
COMPUTER READABLE FORM:

```

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk

```

COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/10/124,557
FILING DATE: 16-Apr-2002
CLASSIFICATION: <Unknown>

CLASSIFICATION: <UNKNOWN>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/643,502
FILING DATE: 18-JAN-1991
APPLICATION NUMBER: US 07/546,114
FILING DATE: 29-JUN-1990
APPLICATION NUMBER: US 07/457,196
FILING DATE: 29-DEC-1989
APPLICATION NUMBER: US 07/390,901
FILING DATE: 08-AUG-1989
ATTORNEY/AGENT INFORMATION:

Query Match 93.2%; Score 5437.9; DB 13; Length 1314;
Best Local Similarity 93.5%; Pred. NO. 4.3e-132;
Matches 1028; Conservative 7; Mismatches 14; Indels 51;

| | | |
|----|---|-----|
| 1 | MAWKTLPIYLLLLLSVFVIOVSSQEL-SCKGRCFSPERGRSCDDDAOCKYDKCCPDY | 59 |
| Qy | | |
| Db | | |
| 1 | MAWKTLPIYLLLSVFVIOVSSQDSSCAGRCGEGSRDATCNDCYNCOHWECCPDF | 60 |
| Qy | | |
| Db | | |
| 60 | ESFCAEVHNPTSPSSKXKAPPSPGASQTIKSTTKRSPKPNKKTKKVISEEITEHSV | 119 |
| Qy | | |
| Db | | |
| 61 | KRVC-----TAHSV | 70 |
| Qy | | |
| Db | | |

22

המחלקה הכלכלית והמנהלית

QY 120 SENQESSSSSIWKIKSSKNSAANRELQALVKVAKNKRKKKPIPKPPVD I79

db 71 SENOESSSSSSSTIWKISSKNISAAANRELOKKLVKDNKKNRTKKKPTPKPPVD 130

.....

0 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31 32 33 34 35 36 37 38 39 40 41 42 43 44 45 46 47 48 49 50 51 52 53 54 55 56 57 58 59 60 61 62 63 64 65 66 67 68 69 70 71 72 73 74 75 76 77 78 79 80 81 82 83 84 85 86 87 88 89 90 91 92 93 94 95 96 97 98 99 100 101 102 103 104 105 106 107 108 109 110 111 112 113 114 115 116 117 118 119 120 121 122 123 124 125 126 127 128 129 130 131 132 133 134 135 136 137 138 139 140 141 142 143 144 145 146 147 148 149 150 151 152 153 154 155 156 157 158 159 160 161 162 163 164 165 166 167 168 169 170 171 172 173 174 175 176 177 178 179 180 181 182 183 184 185 186 187 188 189 190 191 192 193 194 195 196 197 198 199 200 201 202 203 204 205 206 207 208 209 210 211 212 213 214 215 216 217 218 219 220 221 222 223 224 225 226 227 228 229 230 231 232 233 234 235 236 237 238 239 240 241 242 243 244 245 246 247 248 249 250 251 252 253 254 255 256 257 258 259 260 261 262 263 264 265 266 267 268 269 270 271 272 273 274 275 276 277 278 279 280 281 282 283 284 285 286 287 288 289 290 291 292 293 294 295 296 297 298 299 300 301 302 303 304 305 306 307 308 309 310 311 312 313 314 315 316 317 318 319 320 321 322 323 324 325 326 327 328 329 330 331 332 333 334 335 336 337 338 339 340 341 342 343 344 345 346 347 348 349 350 351 352 353 354 355 356 357 358 359 360 361 362 363 364 365 366 367 368 369 370 371 372 373 374 375 376 377 378 379 380 381 382 383 384 385 386 387 388 389 390 391 392 393 394 395 396 397 398 399 400 401 402 403 404 405 406 407 408 409 410 411 412 413 414 415 416 417 418 419 420 421 422 423 424 425 426 427 428 429 430 431 432 433 434 435 436 437 438 439 440 441 442 443 444 445 446 447 448 449 450 451 452 453 454 455 456 457 458 459 460 461 462 463 464 465 466 467 468 469 470 471 472 473 474 475 476 477 478 479 480 481 482 483 484 485 486 487 488 489 490 491 492 493 494 495 496 497 498 499 500 501 502 503 504 505 506 507 508 509 510 511 512 513 514 515 516 517 518 519 520 521 522 523 524 525 526 527 528 529 530 531 532 533 534 535 536 537 538 539 540 541 542 543 544 545 546 547 548 549 550 551 552 553 554 555 556 557 558 559 560 561 562 563 564 565 566 567 568 569 570 571 572 573 574 575 576 577 578 579 580 581 582 583 584 585 586 587 588 589 590 591 592 593 594 595 596 597 598 599 600 601 602 603 604 605 606 607 608 609 610 611 612 613 614 615 616 617 618 619 620 621 622 623 624 625 626 627 628 629 630 631 632 633 634 635 636 637 638 639 640 641 642 643 644 645 646 647 648 649 650 651 652 653 654 655 656 657 658 659 660 661 662 663 664 665 666 667 668 669 670 671 672 673 674 675 676 677 678 679 680 681 682 683 684 685 686 687 688 689 690 691 692 693 694 695 696 697 698 699 700 701 702 703 704 705 706 707 708 709 710 711 712 713 714 715 716 717 718 719 720 721 722 723 724 725 726 727 728 729 730 731 732 733 734 735 736 737 738 739 740 741 742 743 744 745 746 747 748 749 750 751 752 753 754 755 756 757 758 759 760 761 762 763 764 765 766 767 768 769 770 771 772 773 774 775 776 777 778 779 780 781 782 783 784 785 786 787 788 789 790 791 792 793 794 795 796 797 798 799 800 801 802 803 804 805 806 807 808 809 810 811 812 813 814 815 816 817 818 819 820 821 822 823 824 825 826 827 828 829 830 831 832 833 834 835 836 837 838 839 840 841 842 843 844 845 846 847 848 849 850 851 852 853 854 855 856 857 858 859 860 861 862 863 864 865 866 867 868 869 870 871 872 873 874 875 876 877 878 879 880 881 882 883 884 885 886 887 888 889 890 891 892 893 894 895 896 897 898 899 900 901 902 903 904 905 906 907 908 909 910 911 912 913 914 915 916 917 918 919 920 921 922 923 924 925 926 927 928 929 930 931 932 933 934 935 936 937 938 939 940 941 942 943 944 945 946 947 948 949 950 951 952 953 954 955 956 957 958 959 960 961 962 963 964 965 966 967 968 969 970 971 972 973 974 975 976 977 978 979 980 981 982 983 984 985 986 987 988 989 990 991 992 993 994 995 996 997 998 999 1000 1001 1002 1003 1004 1005 1006 1007 1008 1009 1010 1011 1012 1013 1014 1015 1016 1017 1018 1019 1020 1021 1022 1023 1024 1025 1026 1027 1028 1029 1030 1031 1032 1033 1034 1035 1036 1037 1038 1039 1040

QY 180 EAGSLDNGDFRV¹IPD²IS³IQHNKVS⁴SPK⁵IT⁶AKP⁷INFRPSLPFNSDISKEISLIVNK 239

131 BAGSGLDNGDEKVTTPDTS*TOHNKVSTSPKLTAKPINRPSPLPNSDTSKETS*LT*VNK 190

100

1. **Introduction**

QY 240 ETI VEI RE TI I INK QI S I D G K E I I S A K E I Q S I E K I S A N L A P I S K V L A K P I P N A E I I R 299

[illegible]

db 191 ETTVETKETTTNNKOTSTDGKEKTSKAKETOSIEKTSKDLAPTSKVLAKPTPKAETTK 250

100

359

RESULT 14
US-10-124-557-74
; Sequence 74, Application US/10124557
; Publication No. US20020137894A1
; GENERAL INFORMATION:
; APPLICANT: Turner, Katherine
; Clark, Stephen C.
; Jacobs, Kenneth
; Hewick, Rodney M.
; Gesner, Thomas G.
; TITLE OF INVENTION: Megakaryocyte Stimulating Factors
; NUMBER OF SEQUENCES: 143
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genetics Institute, Inc.
; STREET: 87 Cambridgepark Drive
; CITY: Cambridge
; STATE: Massachusetts
; COUNTRY: U.S.A.
; ZIP: 02140
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/10/124,557
; FILING DATE: 16-Apr-2002
; CLASSIFICATION: <unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/643,502
; FILING DATE: 18-JAN-1991
; APPLICATION NUMBER: US 07/546,114
; FILING DATE: 29-JUN-1990

| | | | |
|-----|----|--|-----|
| 241 | QY | TTVETKEITTTNKQTSTDGKEKTTTSKETQSISKTSKAKOLAPTSKVULAKOTPKAEITTKG | 300 |
| 148 | Db | TTVETKEITTTNKQTSTDGKEKTTTSKETQSISKTSKAKOLAPTSKVULAKOTPKAEITTKG | 207 |
| 301 | QY | PALTTPKEPTTPKEBPASTTPKEBPPTTIKSAPTTPKBPATTTKSAPTTPKEBPATTT | 360 |
| 208 | Db | PALTTPKEPTTPKEBPASTTPKEBPPTTIKSAPTTPKBPATTTKSAPTTPKEBPATTT | 267 |
| 361 | QY | KEBAPTTPKEBAPTTPKKEBAPTTPKSAPTTPKEBAPTTPKBPATTTPKBPATTTPKBPPT | 420 |
| 268 | Db | KEBAPTTPKEBAPTTPKKEBAPTTPKSAPTTPKEBAPTTPKBPATTTPKBPATTTPKBPPT | 327 |
| 421 | QY | TTPKBPAPTTPKBPATTPKEBAPTAPKPKBAPTTPKBPATTPKEBAPTTPKESPTTPKE | 480 |
| 328 | Db | TTPKBPAPTTPKEBAPTTPKEBAPTAPKPKBAPTTPKBPATTPKEBAPTTPKESPTTPKE | 387 |
| 481 | QY | PAPTTTTKSAPTTTKBPATTTKSAPTTPKBPSPTTTKEBAPTTPKEBAPTTPKBPATTP | 540 |
| 388 | Db | PAPTTTTKSAPTTTKBPATTTKSAPTTPKBPSPTTTKEBAPTTPKEBAPTTPKBPATTP | 447 |
| 541 | QY | KBPAPTTPKEBAPTTPKKAPATAPKBPAPTTPKETAPTTPKKTTPPTPEKLAPTTPKPA | 600 |
| 448 | Db | KEBAPTTPKEBAPTTPKKAPATAPKBPAPTTPKETAPTTPKKTTPPTPEKLAPTTPKPA | 507 |
| 601 | QY | PTTPPEBLAPTTPEPTTPPEBPAPTTPKAAAPNTPKBPATTPKEBAPTTPKEBAPTTP | 660 |
| 508 | Db | PTTPPEBLAPTTPEPTTPPEBPAPTTPKAAAPNTPKBPATTPKEBAPTTPKEBAPTTP | 567 |
| 661 | QY | KEBAPTTPKGTATTTUKBPAPTTPKBPAPKELAPTTTKBPTSTSTSDKAPTTTPKGTAPT | 720 |
| 568 | Db | KEBAPTTPKGTATTTUKBPAPTTPKBPAPKELAPTTTKBPTSTSTSDKAPTTTPKGTAPT | 627 |
| 721 | QY | PKBPAPTTPKEBAPTTPKGTAPTTTLKEBAPTTPKBPAPKELAPTTTKGPTSTTSDKAPT | 780 |

Db 628 PKPAPPTTPKEPAPPTTPKGTAPTTLUKEPAPPTTPKPAKELAPTTTKGPTSTTSKAPPT 687
Qy 781 TPKEATPTTPKEPAPPTTPKAPPTTPPETPTTSEVSTPTTKEPTTIHKSPDSETPELS 840
Db 688 TPKEATPTTPKEPAPPTTPKAPPTTPPETPTTSEVSTPTTKEPTTIHKSPDSETPELS 747
Qy 841 APTPKALENSPKRGVPTTKTTPAKTKBEMTTAKDITTERDLRTTPTTAAAPKMTET 900
Db 748 APTPKALENSPKRGVPTTKTTPAKTKBEMTTAKDITTERDLRTTPTTAAAPKMTET 807
Qy 901 ATTTEKTTESKITATTQVTSSTTQDTPPFKITTLKTTTLAPKVTITTKKTTITTEIMNKP 960
Db 808 ATTTEKTTESKITATTQVTSSTTQDTPPFKITTLKTTTLAPKVTITTKKTTITTEIMNKP 867
Qy 961 BETAKPKDRATNSKATTPKPKPTKAPKPTSTKPKTTPMVRKPKTTPTRKVTSTMP 1020
Db 868 BETAKPKDRATNSKATTPKPKPTKAPKPTSTKPKTTPMVRKPKTTPTRKVTSTMP 927
Qy 1021 LNPTSRISAEMLQTTTRPNQTPNSKLVEVNPKSEDAGGAGETPHMLLRPHVFMPEVTPD 1080
Db 928 LNPTSRISAEMLQTTTRPNQTPNSKLVEVNPKSEDAGGAGETPHMLLRPHVFMPEVTPD 987
Qy 1081 MDYLPRVNPQGIINPMLS 1099
Db 988 MDYLPRVNPQGIINPMLS 1006

RESULT 15
US-10-124-557-44
; Sequence 44, Application US/10124557
; Publication No. US20020137894A1
; GENERAL INFORMATION:
; APPLICANT: Turner, Katherine C.
; Jacobs, Kenneth
; Hewick, Rodney M.
; Gesner, Thomas G.
; TITLE OF INVENTION: Megakaryocyte Stimulating Factors
; NUMBER OF SEQUENCES: 143
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genetics Institute, Inc.
; STREET: 87 CambridgePark Drive
; CITY: Cambridge
; STATE: Massachusetts
; COUNTRY: U.S.A.
; ZIP: 02140
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/10/124,557
; FILING DATE: 16-Apr-2002
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/643,502
; FILING DATE: 18-JAN-1991
; APPLICATION NUMBER: US 07/546,114
; FILING DATE: 29-JUN-1990
; APPLICATION NUMBER: US 07/457,196
; FILING DATE: 29-DEC-1989
; APPLICATION NUMBER: US 07/390,901
; FILING DATE: 08-AUG-1989
; ATTORNEY/AGENT INFORMATION:
; NAME: Csert, Luann
; REGISTRATION NUMBER: 31,822
; REFERENCE/DOCKET NUMBER: GI 5190
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617)876-1170
; TELEFAX: (617)876-5851
; INFORMATION FOR SEQ ID NO: 44:

; SEQUENCE CHARACTERISTICS:
; LENGTH: 1270 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; SEQUENCE DESCRIPTION: SEQ ID NO: 44:
US-10-124-557-44

Query Match 91.8%; Score 5353.7; DB 13; Length 1270;
Best Local Similarity 91.5%; Pred. No. 66-130;
Matches 1006; Conservative 0; Mismatches 0; Indels 93; Gaps 1;

Qy 1 MAWKTLPIYLLLLSVFVIOQVSSOELSCGKRCFBSFERGECDCDAQCKYDKCCPDYE 60
Db 1 MAWKTLPIYLLLLSVFVIOQVSSOELSCGKRCFBSFERGECDCDAQCKYDKCCPDYE 60
Qy 61 SFCAEVHNTSPSPSSKKAPPPSGASQTIKSTTKRSFKPKPKKTKKXVIESEITEHSVS 120
Db 61 SFCAE----- 65
Qy 121 ENQESSSSSSSSSSSTIWKIKSSKNSAANRELQKKLVKONKKKRTKKKTPKPPVDE 180
Db 66 -----VKONKKKRTKKKTPKPPVDE 87
Qy 181 AGSGLDNGDFKVTTPDTSTTOHNVKSTSPKITTAKPINRPSLPSPNSDTSKETSITVKE 240
Db 88 AGSGLDNGDFKVTTPDTSTTOHNVKSTSPKITTAKPINRPSLPSPNSDTSKETSITVKE 147
Qy 241 TTVEIKETITTKQSTDGKETTSAKETOSIEKTSADLAPTSKVLAKPTKAEITTKG 300
Db 148 TTVEIKETITTKQSTDGKETTSAKETOSIEKTSADLAPTSKVLAKPTKAEITTKG 207
Qy 301 PALTTPKETPTTPKEPASTTTPKEPTTIKSAPTTPKEPAPTTTKSAPTTKSAPTTTKEPAPTTT 360
Db 208 PALTTPKETPTTPKEPASTTTPKEPTTIKSAPTTPKEPAPTTTKSAPTTKSAPTTTKEPAPTTT 267
Qy 361 KGPAPTTPKPAPTTTKGPAPTTTKSAPTTPKEPAPTTTKGPAPTTTKGPAPTTTKGPAPTT 420
Db 268 KGPAPTTPKPAPTTTKGPAPTTTKSAPTTPKEPAPTTTKGPAPTTTKGPAPTTTKGPAPTT 327
Qy 421 TTPKEPAPTTKEPAPTTTPKEPAPTTAPKAPPTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKE 480
Db 328 TTPKEPAPTTKEPAPTTTPKEPAPTTAPKAPPTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKE 387
Qy 481 PAPTTPKSAPTTTKEPAPTTTKSAPTTPKPSPTTKGPAPTTTPKEPAPTTTPKEPAPTTTP 540
Db 388 PAPTTPKSAPTTTKEPAPTTTKSAPTTPKPSPTTKGPAPTTTPKEPAPTTTPKEPAPTTTP 447
Qy 541 KGPAPTTPKPAPTTTKGPAPTTTKGPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKE 600
Db 448 KGPAPTTPKPAPTTTKGPAPTTTKGPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKE 507
Qy 601 PTTPEELAPTTPEEPTTPPTPEPAPTTTPKAAAPNTPKPAPTTTPKEPAPTTTPKEPAPTTTP 660
Db 508 PTTPEELAPTTPEEPTTPPTPEPAPTTTPKAAAPNTPKPAPTTTPKEPAPTTTPKEPAPTTTP 567
Qy 661 KETAPTTKGTAPTTLKGPAPTTTPKGPAPTKELAPTTTKEPTSTTSKAPTTKGTAPTT 720
Db 568 KETAPTTKGTAPTTLKGPAPTTTPKGPAPTKELAPTTTKEPTSTTSKAPTTKGTAPTT 627
Qy 721 PKPAPTTTPKEPAPTTTPKGTAPTTLKGPAPTTTPKGPAPTKELAPTTTKEPTSTTSKAPTT 780
Db 628 PKPAPTTTPKEPAPTTTPKGTAPTTLKGPAPTTTPKGPAPTKELAPTTTKEPTSTTSKAPTT 687
Qy 781 TPKEATPTTPKEPAPTTTPKAPPTTPPETPTTSEVSTPTTKEPTTIHKSPDSETPELS 840
Db 688 TPKEATPTTPKEPAPTTTPKAPPTTPPETPTTSEVSTPTTKEPTTIHKSPDSETPELS 747
Qy 841 APTPKALENSPKRGVPTTKTTPAKTKBEMTTAKDITTERDLRTTPTTAAAPKMTET 900
Db 748 APTPKALENSPKRGVPTTKTTPAKTKBEMTTAKDITTERDLRTTPTTAAAPKMTET 807
Qy 901 ATTTEKTTESKITATTQVTSSTTQDTPPFKITTLKTTTLAPKVTITTKKTTITTEIMNKP 960

Db 808 ATTTEKTESKITATTQVTSITTQDTPFKITTLKTTTLAPKVTTTKKTTITTEINNKP 867
QY 961 EETAKPKDPAATNSKATTPKPKPTKAPKPTSTKKPKTTPRVRKPKTTTPRKYTSTMP 1020
Db 868 EETAKPKDPAATNSKATTPKPKPTKAPKPTSTKKPKTTPRVRKPKTTTPRKYTSTMP 927
QY 1021 LNPTSRIAEAMLQTTTRNQTSNSKLVEVNPKSEDDAGGAEGETPHMLLRPHVFMPEVTPD 1080
Db 928 LNPTSRIAEAMLQTTTRNQTSNSKLVEVNPKSEDDAGGAEGETPHMLLRPHVFMPEVTPD 987
QY 1091 MDYLPVPVNPQGIINPMLS 1099
Db 988 MDYLPVPVNPQGIINPMLS 1006

Search completed: October 13, 2004, 11:52:38
Job time : 117.81 secs

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: October 13, 2004, 11:23:49 ; Search time 22.3855 Seconds
(without alignments)
4723.689 Million cell updates/sec

Title: SEQ1-A
Perfect score: 5835
Sequence: 1 NAWKTLPIYLLLLSVFVIQ.....DMDYLPVNPQGIINPMLS 1099

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.1

Searched: 283416 seqs, 96216763 residues

Total number of hits satisfying chosen parameters: 283416

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

PIR79:*
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | ID | Description |
|------------|--------|-------------|--------|--------|--------------------|
| 1 | 1283.7 | 22.0 | 3020 | A43932 | mucin 2 precursor, |
| 2 | 1143.3 | 19.6 | 1664 | T18262 | S-layer protein - |
| 3 | 1015.5 | 17.4 | 1274 | T16251 | hypothetical prote |
| 4 | 1014.9 | 17.4 | 2187 | T30826 | nascent polypeptid |
| 5 | 1009.9 | 17.3 | 1489 | T31108 | cyst germination s |
| 6 | 1008.4 | 17.3 | 7862 | T38346 | elastic titin - hu |
| 7 | 994.8 | 17.0 | 3570 | T45025 | mucin MUC5B, trach |
| 8 | 933.8 | 16.0 | 1367 | S48478 | glucan 1,4-alpha-g |
| 9 | 896.2 | 15.4 | 3507 | T34513 | hypothetical prote |
| 10 | 886.7 | 15.2 | 1188 | S49915 | extensin-like prot |
| 11 | 879.9 | 15.1 | 6642 | T23757 | protein UNC-99 - C |
| 12 | 875.8 | 15.0 | 2897 | T48666 | cell proliferation |
| 13 | 871.4 | 14.9 | 1229 | T25697 | hypothetical prote |
| 14 | 863.5 | 14.8 | 3256 | A48666 | cell proliferation |
| 15 | 860.7 | 14.8 | 1151 | T18535 | high molecular mas |
| 16 | 848.1 | 14.5 | 5762 | A1819 | proline-rich pepti |
| 17 | 837.8 | 14.4 | 1344 | A35175 | mucin 1 precursor, |
| 18 | 794.4 | 13.6 | 4135 | T43629 | tenascin-X - bovin |
| 19 | 794.1 | 13.6 | 2332 | T34434 | hypothetical prote |
| 20 | 792.1 | 13.6 | 2142 | T35098 | MHC class III hist |
| 21 | 785.9 | 13.5 | 1832 | T31113 | mucin-like glycopo |
| 22 | 777.7 | 13.3 | 990 | IS1618 | nucleolar phosphop |
| 23 | 775.7 | 13.3 | 3942 | T42730 | Bassoon protein - |
| 24 | 773.2 | 13.3 | 2774 | A43359 | microtubule-associ |
| 25 | 766 | 13.1 | 1872 | S36152 | MHC class III hist |
| 26 | 760.2 | 13.0 | 1870 | S37671 | MHC class III hist |
| 27 | 756.1 | 13.0 | 761 | C84672 | hypothetical prote |
| 28 | 751.6 | 12.9 | 4548 | S00657 | apoptosis(a) (EC |
| 29 | 748.4 | 12.8 | 5362 | T03454 | ALR protein - huma |

ALIGNMENTS

RESULT 1

A43932 mucin 2 precursor, intestinal - human (fragments)

N;Alternate names: mucin SMUC-41

C;Species: Homo sapiens (man)

C;Date: 10-Mar-1993 #sequence revision 12-Apr-1996 #text change 09-Jul-2004

C;Accession: A49963; A45106; B45106; A43932; B33532; A61257; PQ0328; PQ0329

R;Gum J.R.; J.R.; Hicks, J.W.; Toribara, N.W.; Siddiki, B.; Kim, Y.S.

J. Biol. Chem. 269, 2440-2446, 1994

A;Title: Molecular cloning of human intestinal mucin (MUC2) cDNA. Identification of the

A;Reference number: A49963; MUID:94132002; PMID:8300571

A;Accession: A49963

A;Molecule type: mRNA

A;Residues: 1-639 <GUL>

A;Cross-references: UNIPROT:Q02817; UNIPROT:Q14884; GB:L21998

R;Gum J.R.; J.R.; Hicks, J.W.; Toribara, N.W.; Rothe, E.M.; Lagace, R.E.; Kim, Y.S.

J. Biol. Chem. 267, 21375-21383, 1992

A;Title: The human MUC2 intestinal mucin has cysteine-rich subdomains located both upstr

A;Reference number: A45106; MUID:93016075; PMID:1400449

A;Accession: A45106

A;Status: not compared with conceptual translation

A;Molecule type: mRNA

A;Residues: 626-1895 <GU2>

A;Cross-references: GB:M94131; NID:G186395; PIDN:AAA59163.1; PID:G186396

A;Note: sequence extracted from NCBI backbone (NCBIP:116706)

A;Accession: B45106

A;Status: not compared with conceptual translation

A;Molecule type: mRNA

A;Residues: 2037-3020 <GU3>

A;Cross-references: GB:M94132; NID:G186397; PIDN:AAA59164.1; PID:G186398

A;Experimental source: colon

A;Note: sequence extracted from NCBI backbone (NCBIP:116698)

R;Toribara, N.W.; Gum Jr., J.R.; Culhane, P.J.; Lagace, R.E.; Hicks, J.W.; Petersen, G.M.

J. Clin. Invest. 88, 1005-1013, 1991

A;Title: MUC-2 human small intestinal mucin gene structure. Repeated arrays and polymorp

A;Reference number: A43932; MUID:91358717; PMID:1885763

A;Accession: A43932

A;Molecule type: DNA

A;Residues: 1343-1350, 'L', 1352-1411, 'S', 1413-1448, 'P', 1450-1503, 'T', 1505-1915 <TOR>

A;Cross-references: GB:M74027; NID:G188863; PIDN:AAA59875.1; PID:G188864

A;Note: sequence inconsistent with the nucleotide translation

A;Note: sequence extracted from NCBI backbone (NCBIP:55749, NCBIP:55750)

R;Gum, J.R.; Byrd, J.C.; Hicks, J.W.; Toribara, N.W.; Lamport, D.T.A.; Kim, Y.S.

J. Biol. Chem. 264, 6480-6487, 1989

A;Title: Molecular cloning of human intestinal mucin cDNAs. Sequence analysis and eviden

A;Reference number: A33532; MUID:89197956; PMID:2703501

A;Accession: B33532

A;Molecule type: mRNA

A;Residues: 1916-2193 <GU4>

A;Cross-references: GB:M22405; NID:G188873; PIDN:AAA63334.1; PID:G188874

A;Experimental source: intestine

R;Jany, B.H.; Gallup, M.W.; Yan, P.S.; Gum, J.R.; Kim, Y.S.; Basbaum, C.B.

J. Clin. Invest. 87, 77-82, 1991
 A:Title: Human bronchus and intestine express the same mucin gene.
 A:Reference number: A61257; PMID:91086481; PMID:1985113
 A:Accession: A61257
 A>Status: not compared with conceptual translation
 A:Molecule type: mRNA
 A:Residues: 71, 1925-1948, 'TTS', 1952-1954 <JAN>
 A:Experimental source: bronchus
 R:Xu, G.; Huan, L.; Khatri, I.; Sajjan, U.S.; McCool, D.; Jones, C.; Forstner, Biochem. Biophys. Res. Commun. 183, 821-828, 1992
 A:Title: Human intestinal mucin-like protein (MLP) is homologous with rat MLP in the C-terminus
 A:Reference number: P00328; PMID:192198477; PMID:1550588
 A:Accession: P00328
 A:Molecule type: mRNA
 A:Residues: 2328-2468 <XUG>
 A:Cross-references: GB:M85523
 A:Experimental source: small intestine
 A:Accession: P00329
 A:Molecule type: protein
 A:Residues: 2328-2342, 'K', 2344-2354 <XUG1>
 C:Genetics:
 A:Gene: GDB:MUC2
 A:Cross-references: GDB:120203; OMIM:158370
 A:Map position: 11p15.5-11p15.5
 C:Superfamily: von Willebrand factor; von Willebrand factor type A repeat homology; von Willebrand factor type C repeat homology
 C:Keywords: glycoprotein; intestine; tandem repeat
 F:2765-2834/Domain: von Willebrand factor type C repeat homology <VMC>
 Query Match 22.0%; Score 1283.7; DB 2; Length 3020;
 Best Local Similarity 21.5%; Pred No. 6.3e-18;
 Matches 406; Conservative 97; Mismatches 498; Indels 883; Gaps 61;
 3 WK-----TLPIYLLLSVFIQVSSQELSCKGRCFESF-----E 38
 649 WREHVCKNDVSCPNQVFLNLTTCQTCRSLSEADSHCLGEPFAPVDGCGCPDHTLDE 708
 39 RGR-----ECDC-----DAOCKYDXC-CPDIYESFCAEVH----- 67
 709 KGRVPLAKSCYHRLGYLEAGDVVVRQBERCVCRDLHCRQLRLIGQSCTAPKIMDC 768
 68 -NPTSPSSKAPPPSGASQTIKS-----TTKSPKPPN 100
 769 SNLTALATSK---PRALSCQILAAGYHTECVSGCVCPDGLMDGRCGVVEKECP--- 821
 101 KKTKKVIESBEITEHVSSENOESSSSSSSSSTTWIKSKKNAA----- 149
 822 -----CVHNDLYSSGA-----KLVDCNTCTCKRGRWVCTQA 854
 150 -----NREL----- 153
 855 VCHGTCISYSGHYITFGKYYPDGHCSYAVQDYCGQSSLSGSFSLITENVPCGTTGV 914
 154 -----QKLVKVDNKK----- 167
 915 TCSKAIKIFMGRTLEKDEKRWVQIRDEGHVAYTTREVGVLYVVSSTGIIIVINDKRT 974
 168 ---KKKTPKPPVVDAGSL-----DNGDFKV----- 196
 975 TVFKLAFSYKGV-----CGLGNFDRHSNNDFTTRDHVVVSSELDGNSWEAPTCPD 1029
 197 TSTTQ-----HNKYSTSP----- 209
 1030 VSTNPEPCSLNPHRRSWEAKQCSILKSVFISCHSKVDPKPFYACVHDSGCDTGDC 1089
 210 -----KLTAKPIN 218
 1090 CFCGAVASAOECKEGACVFWRTDLCIPFCDYNNPHECEWHYPCGNRSFETCRTIN 1149
 219 -----PPSLP----- 224
 1150 GIHSNISVLEGVCYPCPKDRPIVEEDLKKCVTADKGCYVEDTHYPPGASVPTETCK 1209
 225 -----PNSDTSKE-----TSLIVNKET----- 241

Db 1210 SCVCTNSSQVCPBEGKILNQTDGAPCYWEICGNGVGVKHFNICSITTPSLTITFT 1269
 QY 242 -----TVETKETTNTNKQSTIDGKERTTSK----- 267
 Db 1270 TITLPTPTPTSTTTTTTTTTSTVLSLTPKLCCLMSWDINEDHPSSGSDGDRPFQGV 1329
 QY 268 --EQSISTKTSKD----- 279
 Db 1330 CGAPEIDERSVXDPHLSUEHQGQKQCVDSVGFICKNEDQFGNGPFGLCYDYKIRVNCC 1389
 QY 280 -----LAPTSKVLAKPTPKAETTTKGPALTTTKEPTTTPKEPASTTTKEPTTTIKSA 333
 Db 1390 WPMDCIITTPSPPTTTPSPPTTTPPTTTPSPPTTTPPTTTPSPPTTTPPTTTP 1449
 QY 334 P-TTPKEPAPTTTKSAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTP 391
 Db 1450 PTTTPSPPTTTPPTTTPSP-PPTTTPSPPTTTPSPPTTTPPTTTPSPPTTTP 1507
 QY 392 -----KEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAP 443
 Db 1508 ITPPASTTTLPTTTPSPPTTTPPTTTPSPPTTTPPTTTPPTTTPPTTTPPTTTP 1567
 QY 444 TAPKAPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTP 503
 Db 1568 TTTTPPTTTPSPPTTTPSPPTTTPPTTTPSPPTTTPPTTTPSPPTTTPPTTTPSP 1624
 QY 504 APPTKEPSPPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTT 562
 Db 1625 -PTTP-----PTSTTLPTTTPSPPTTTPPTTTPSPPTTTPPTTTPSPPTTTPPTT 1679
 QY 563 APKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTP 622
 Db 1680 TPSSPTTTP--SPPTT-----TWTTTPS--PTTTPSSPTTTPPTTTPSPPTTTP 1730
 QY 623 PAPTTPKAAAPNTKEPA-----PTTPKEPAPTTTPKEPAPTTTP--KETATTP-- 668
 Db 1731 PSPTT-----TPSPPTTTPPTTTPPTTTPPTTTPPTTTPPTTTPPTTTPPTTTP 1783
 QY 669 -----KG----- 670
 Db 1784 PLCNWTGLDSCGNPKFPGGDTLIGDVGCPGWAANI-SCRATMYPDVPIGLQGVVCD 1843
 QY 671 -----TAPTILKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTT 694
 Db 1844 VSVGLICKNEDQKQGVIPMAFLNYEVQCECVTQPTM---TTTTENPTPTPTT 1900
 QY 695 TT--TKPEPTSTSDKP---APT-TPKGTAPTTPKEPAPTTTPKEPAPTTTPKEPAPTT 747
 Db 1901 TTTVTPPTPTSTQSPNGLOAPTPTPTSTTTVTPPTPTPTPTPTPTPTPTPTPTPTPT 1959
 QY 748 PAPTTPKAPKAPKAPKAPKAPKAPKAPKAPKAPKAPKAPKAPKAPKAPKAPKAPKAP 801
 Db 1960 PPTPTGQTPTTVLITTTTTTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPT 2019
 QY 802 --APTTPPT 854
 Db 2020 STTTTPT 2079
 QY 855 PGVPTTKTAAKPEMTTAKDPTTERDLRTTPTTTAAKPMKETAATTTKTTESKITA 914
 Db 2080 STLPALIENTSTAPPSTPTPTPTTSGHTLSPPTTTPPTTTPPTTTPPTTTPPTTTP 2137
 QY 915 TTTQVTSITTTQDTPKLTIT---LKTTLAPKVVITTKTITITTEIMNKEE----- 962
 Db 2138 STVQITTTSAWPT 2197
 QY 963 -----TAKPKDRATNSKATTPKPKPTKAPKPTSTKPKPTKPTSTKPKPTKPT 1002
 Db 2198 TCVFVNCSLSCLEFYNWNSCPSTPTPTPSK-SPTPSKSPSTPSKPTPTPTKPECPDF 2256
 QY 1003 RPKPTTTPKMTSTMPBLNPTSRITAEAMLOTTTPRQNTNSKLVEVNP----- 1051

Db 2257 DDPK-----QENETWLCDFM-ATCKYNNVTEIVKVECEPPMPTCSNGLQ 2302
Qy 1052 --KSDAGGA-----EGTPEMLLRHVFMEVETPDM 1081
Db 2303 PVKVEDPDCCHWECDCYCTGWDPHYVTFDGLYSYQGNCTYLVLE-----EISPSV 2356
Qy 1082 D-----1082
Db 2357 DNFGVYIDNYHCDPNDKSCPRTLIVRHETQEVLIKTVMMPMQVQVQNRQAVLPYKK 2416
Qy 1083 -----YLPVFNQGIIN 1095
Db 2417 YGLEVYQSGINVVDPILGVLVS 2440
RESULT 2
T18262
S-layer protein - Clostridium thermocellum
C:Species: Clostridium thermocellum
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 09-Jul-2004
C:Accession: T18262
R:Fujino, T.; Beguin, P.; Aubert, J.P.
J. Bacteriol. 175, 1891-1899, 1993
A:Title: Organization of a Clostridium thermocellum gene cluster encoding the cellulose
e.
A:Reference number: Z18847; MUID:93209931; PMID:8458832
A:Accession: T18262
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-1664 <FUJ>
A:Cross-references: UNIPROT:Q06852; EMBL:X67506; NID:9256879; PID:9296881; PIDN:CAA47841
Query Match 19.6%; Score 1143.3; DB 2; Length 1664;
Best Local Similarity 21.2%; Pred. No. 1.6e-15;
Matches 367; Conservative 131; Mismatches 425; Indels 807; Gaps 66;
Qy 4 KTLPIYLLLLSVFV-----IQOVSSQE 26
Db 6 KVLISLLTLLTISTTSVNMSPAEATPSIEMVLDKTEVHGVITATIKVNNIRLAGYQ 65
Qy 27 LSCGKRCFSFERGECDCQAQCKYDKCPDYSCAEVHNFTSPFSSK-----KAPPPS 82
Db 66 LNIK-----FDPEVLQVPDPCATGERTKSMF-- 92
Qy 83 GASQIKSTTKSPPPNKKTKKVIIESEITEHSVSENQSSSSSSSSSSSTIWKIK 142
Db 93 -VNRVLLTNSKYGTP-----VAGNDIKSGIINFATGYNLLTAYK 131
Qy 143 SS-----PKPPVDEAGSLDN-----187
Db 132 SSGIDEHTGIIIGEIKFVLKKONTSIRFEDTSLMFGAISGTSLFDWDAETITGYEVIQPD 191
Qy 145 -----KNSAANRELOK-KLKYKD-----NKKQ-----RTKKKPT-----172
Db 192 LIWAEAPLKDASVALELDKTKVKYGDIIITATIKIENKMFAGYQYLNKYDPTMLEAIEL 251
Qy 173 -----PKPPVDEAGSLDN-----187
Db 252 ETGSAIAKTPWVGTVLQSNYKTTAVANDVGAGIINFABEYNSLTKYRETGVABET 311
Qy 188 -----GDFKVTTPDT-----STQHNKV-----205
Db 312 GIIGKIGFRVLKAGSTAIRFEDTAMPGAIEGYMFDWYGENIKGYSVVQGEIYAEGEE 371
Qy 206 -----STSPKITTAKINRPSLP-----PNS-----DTSK-----231
Db 372 PGEETEEFVPTETPVDPPTVTEBFVPSELPSVIMELDKTKVKVKGVDIIITATIKIENM 431
Qy 232 -----ETSLVNKETTETKETTINKQTSIDGKEKTS-----265
Db 432 KNPAGYQLNIKVDPTMLEAIELETGSAIAKRTWPVTGGTV-----LQSDNYKTTAVAND 486
Qy 266 -----AKETQSIKTSKOLAPTSKVLAKPTPKAETTT-----298

Db 487 VGAGIINFABEYNSLTKYRETGVABETGIIIGKIFRVLKAGSTAI-----RPEDTTAMPG 541
Qy 299 -----KGPALTTT-----KBPPTT-TPKEPASTTPKEPPTT---328
Db 542 AIEGTVMFDWYGENIKGYSVVQGEIYAEGEETEBPPTETFDPTFTVTEBVPSEL 601
Qy 329 -----328
Db 602 DSYVIMELDKTKVKEGDVLIATIRVNNIKNLASQIGIKYDPAKLEAFNIETGDPIDEGT 661
Qy 329 -----T 329
Db 662 WPAVGGTILKNRDYLPITGVAINNVSKGILNFAAYVYVFDYBEGKSEDTGIIIGNIFRV 721
Qy 330 IKSAPTTPK-----EPATTTKSAFTTKBPAPT 359
Db 722 LKAEDTIRFEELESMPGSDGYMLDYNLRISGVVVIQPAIKAA-----DEIPTD 776
Qy 360 TKEPAPTTKKEPAP-----TTTKBPAPTTKSAFTTKBPAPTTPKBPAPTTPKEPAPT--413
Db 777 TPSDEPTSPDEPTSPDEPTSPDEPTSPDEPTSPDEPTSPDEPTSPDEPTSPDEPTSPDE 836
Qy 414 -----TPK-EPPTT-TPKEPAPT-TKEPAPTTKBPAPTAPKKA-----PTTKBPAPT-460
Db 837 PTPSDEPTSPDEPTSPDEPTSPDEPTSPDEPTSPDEPTSPDEPTSPDEPTSPDEPTSP 896
Qy 461 TPKEPAPTTTKBPAPTTPKBPAPTTPKBPAPTTPKBPAPTTPKBPAPTTPKBPAPTTPK 520
Db 897 TPEPIPTDTPSDEPTSPDEPTSP-----DEPTSPDEPTSPDEPTSPDEPTSPDEPTSP 953
Qy 521 PTPKBPAPTTPKKA-----PTTKBPAPT-TPKEPAPTTTKBPAPTTPKBPAPTTPK 565
Db 954 PTPSDEPTSPDEPTSPDEPTSPDEPTSPDEPTSPDEPTSPDEPTSPDEPTSPDEPTSP 1013
Qy 566 EPAPTTKBPAPTTPKKTTPKKTTPKKTTPKKTTPKKTTPKKTTPKKTTPKKTTPKKTTP 621
Db 1014 EPTPSD-----EPTPSDDEPTSPDEPTSPDEPTSPDEPTSPDEPTSPDEPTSPDE 1052
Qy 622 EPAPTTKKAAPNTKBPAPTTPKKA-----EPTTKBPAPT-TPKETAPTTKBPAPTTPKE 678
Db 1053 EPIPTDTPSDEPTSPDEPTSPDEPTSPDEPTSPDEPTSPDEPTSPDEPTSPDEPTSPDE 1112
Qy 679 PAPTTPKBPAPKELAPTTKEPT-SKTSKBPAPTTPKGTAPTTKBPAPTTPKBPAPTTPK 737
Db 1113 PTPS--DEPTSPD-EPTSPDEPTSPDEPTSPDEPTSPDEPTSPDEPTSPDEPTSP--1165
Qy 738 KGTAAPTLKBPAPT-TPKBPAPKELAPTTKGTSTTSDKPAPT---TPKETAPTTKBP 793
Db 1166 --DEPTSPDEPTSPDEPTSPDEPTSPDEPTSPDEPTSPDEPTSPD-EPTSPDEP 1217
Qy 794 APT-TPKBPAPTTPPTTPPTSPPTSPPTTPKBTTHKSPDEPTSPDEPTSPDEPTSPDEPT 852
Db 1218 TPSETPEPIPTDTPSDEPTSPD-EPTSPDEPTSPD-EPTSPDEPTSPD-EPTSPD--SETP 1266
Qy 853 KEGVPTTKTAPKAPKEMTTAKDKTTERDLRTTPETTTAAPKMTKETAATTEKTTESKI 912
Db 1267 EEP-IPTDTPSDEPTSPDEPTSPDEPTSPDEPTSPDEPTSPDEPTSPDEPTSPDEPTSP 1325
Qy 913 TATTTQVTSITQDTPFKITTLTKTTLAPKVTTKTITTTIMKNKEETAKPKRATN 972
Db 1326 PTDTPSDEPTSPDEPTSPDEPTSP-----PSDEPTSPDEPTSP 1360
Qy 973 SKATTPKQ-KPKAPKAPTSTKKPKTMRVRPKTTPTPKMTS-----1016
Db 1361 SDEPTSPDEPTSPDEPTSP-----PTTPTTPTTPTTPTTPTTPTTPTTPTTPTTPTTPT 1407
Qy 1017 -----TWPELNPTSRIAEAMLQTTTRNPQTPNSKLVFNPKSEDAAGAGETPHMLL 1068
Db 1408 GGGGGTVPSTPTPTPTS-----KPTSTAPTEIE-EPTPSDVPAIGGEHAYL 1455
Qy 1069 RPH--VFME-----1076

A:Molecule type: DNA
A:Residues: 1-2187 <YOT>
A:Cross-references: UNIPROT:P70670; EMBL:U48363; NID:G1666688; PID:G1666689; PIDN:AAB187
C:Gene: Naca
A:Map position: 10
A:Introns: 24/1; 1996/1; 2050/3; 2059/3; 2142/3; 2183/3
A:Note: differential splicing converts alphanac into a tissue-specific DNA-binding active
C:Keywords: alternative splicing; DNA binding; transcription factor

Query Match 17.4%; Score 1014.9; DB 2; Length 2187;
Best Local Similarity 19.3%; Pred. No. 7.9e-13;
Matches 375; Conservative 147; Mismatches 426; Indels 991; Gaps 73;

QY 21 QVSSQ-ELSCGRCFESPERGECDCDAQCKKYDKCCPD-YESFCAEVHNP----- 69
DB 183 QVPSQGTNLKGT-----PCPDVVPAPFPHLENPLASVQPGLM 222
QY 70 -----TSPSSKKAPP----- 80
DB 223 SCPTLSNTSPKGVPISSALTQSRLSLNLKGVSPPARNTAAPSILAPSTSLGCHLPL 282
QY 81 ----- 80
DB 283 LHHSSVDSPIQPGQSGGLAVNPTSVGHGIAASCPCPCVUPALPSRLLAVDGAAPSD 342
QY 81 -----PGASQTIKSTTKRSPK-----PNKKTKKVI 108
DB 343 DKGSAVNLCSPPGSSNV--AGTSLSPKASLVKPGSNVALQPLVTVQVPAKQGLKEI 400
QY 109 ESEETEEHVSSEN----- 122
DB 401 PVSCIGATHALDNPSAISVAPATHVPPPTSSGLVSKDPAFVTSLVVPAAHKQFPAPP 460
QY 123 -----QSSS 126
DB 461 ASATLGVFVPLPATEGLKLNLPISALVNVGAPVSPAQAGLPTRKDTLLQPLAPIALKESP 520
QY 127 SSSSSS--SSSTIWK----- 140
DB 521 SSSQSSSLEVLSEDVTKTTGGPAPVVRPAIAGVATTSLRADSPFAVIRADSCVSPNT 580
QY 141 -----IKSKNSAANR----- 151
DB 581 VSQPLKRSVTDPAAPRAKNTAPSTTSLVPLASEGCPVASSMALSQNASVSETALAL 640
QY 152 --ELQKLVKDNKXNRKKKTPKPPV-----VDEAGSLDNGDFKVTTPDTST 199
DB 641 SPEIPKSV-----PFPDPLAEISFSNARKVDVSHMESGSSRQGHPDASV 687
QY 200 TQH-----NKYST 207
DB 688 TAKGTVVCLADSSLDTSVSASGSLASGASSPLFVFSFUEAGLVGQPKGSLNKLSP 747
QY 208 SPKITTAKPINRPSLPPNSDTSKETSNTVNKETTVEKE----- 247
DB 748 TFPSSKGAIV-FSTGAPP---SPKGAPIVPTSEISSKQVPAEILPSPQKTPVETASRL 802
QY 248 -----TTTNKQTDGKEKITSKETSIEKTSQ-----KDLAPT 283
DB 803 ISAVOSPKVDPIMSDVPTSPKXTSATAPKDTSA--TLSLSKVPAVTSLSPPKAPVAPS 860
QY 284 SKVLAKP-----TPKAEITTKGPALT----- 305
DB 861 NEATIVPTIELTSLNALAAATPKETLATSIPKVTSPSPQKTPKSVSLKGAPAMTSKAT 920
QY 306 -----PKTTPKEPASTT-----PKETPTTKKAPT 336
DB 921 ELAASKDVSPSPKPEVPLQHVPTSPKSPVSDTLGALTSPPPKGP-PATLAETPTY 979
QY 337 PKE-PAPTTTKSAPTPK-----EPAPTTKEBAPTPPKBAPTTKEPAP----- 381
DB 980 PKKSPKPAASKKTPATPSPGVTAVPLEIPCKSKKAPKTAAPKESATSSSKRAPKTAVS 1039

QY 382 -----TTTKSAPTTPKBAPTTPKK-PAPTPKE----- 409
DB 1040 KEIPSKGVTAIVLEISLPLKETSASA--TPGEKASSPKRSKPTAGKETPPGGVTAIVPP 1097
QY 410 -----PAPTTKEPTP-----TTTKE-----PAPTTKE 432
DB 1098 EISLPPKRETONATPNESLAASQKRSKPTSVKETPPGGVTAMPLEIPSAQKAPKTAV 1157
QY 433 P-----APTTKSPAPTAKKPAPTP----- 454
DB 1158 PKQIPTPEDAVTILAGSPLSPKASKTAAPKEAPATPSVGVIAVSEISPSPKKTSKATA 1217
QY 455 -----KBPAPTTKEPAPTTTKE-----PSPTPKE---PAPTTKSAP 490
DB 1218 PKENSATLPPKRSKTAAPKE--TPATSSSEGVTAVPSEISPSPTPASKGVVLTLPKGAP 1276
QY 491 TTTKEBAPTTTKSAPTT--PKES--PTTKBAPTTPKEBAPTTPKKKAPTT----- 539
DB 1277 NALAE-SPASPKVKPTAAPEETSTTSPQIKFKVAGPKESATSPSKKTPKTAIVPKETS 1335
QY 540 -PKBAPTTPKEBAPTTPKKAPTAAPKE--PAPTTPKETAPTTPKKLTPTTPEKLAPTPE 597
DB 1336 APSEGVTAIVLEIPSPSKAPKTAAPKETAPS--PEGATTAPVQI--PPSPRKGSKKAGS 1392
QY 598 KPAPTTPEELAPTTPEPTTPEEP--APTTPKAAAP-----NTP 636
DB 1393 KETPTTP-----SPEGVTAAPLEIPTSCKKTSKMASPKETLVTPSSKLSQTVGPKETS 1446
QY 637 KEBAPTTPKEBAPTTPKEBAPTTPKETAAPTTPKGTAPTTLKEBAPTTPKKAPKELAP-- 694
DB 1447 LEGATAVLEIPSHKAPKTVDPKQVLTTPSKDAPTTLAE--SPSPKK-APKTAAPPS 1504
QY 695 --TTTKEPTSTSDKPAITTPKGTAPTTPKBAPTTPKEPA-----PTTPKGTAPT 743
DB 1505 ERVTIVFP-----EKPA--TPQKASGTTASKVPVPAETQEVAVSSRETFVTPVAVPVKNPS 1558
QY 744 TLKEBAPTTPKKAPKELAPTTKGTSTSDPAPTTPKETAAPTTPKEBAPTTPKKAP 803
DB 1559 SHKTSKTIELKCAPATLPSPTKSPKIPSSKK-----APTSAP-----KEFP 1602
QY 804 TTPTETPPTTSEVSTPTTKEPTTIHKSPDESTPE--LSAETPKALENSKPEPGVPTTK 861
DB 1603 ASFSIKVTTSLAQ-----APSLQKAPSTTIPKENLAA--PAVLVSSKSPAAP-AR 1653
QY 862 TPAATKEEMTTAKDKTTERDLRTP-----ETTTA-----APKMTKETATTT 904
DB 1654 ASASLSP---ATAAPQAPKEATTIPCKKAAATETPIETSTAPSLEGAPKETSETSV-- 1708
QY 905 EKITESKITATTQVTSITTTQDTPPKITL---KTTTLAPKVTTTKTITTEIMNKP 960
DB 1709 -----SKVLMSSPPKASSSKRASTLPATTLPSLKEASVLSPTATSSGK----- 1752
QY 961 EETAKPKDRATNSKATTPKQKPTKPKK----- 989
DB 1753 DSHISPVSDACSTGTTT--PQASEKLPSKKGPTAFTMLAAPAESAALAITAPIQKSPGA 1810
QY 990 -----PTSTKPKTMR--VRKPKTTTTPKMTSTWPELNPTSEIAEMLOTT 1035
DB 1811 NSNSASSPKCPDSSSKDRTKGLPSAVALAPQTVPEK-----DTSKALETLVSP 1860
QY 1036 TRNQ-----TPNSKLV 1047
DB 1861 AKGSDCLHSPKGVGSOVATFLAAFTSDKVPPEAVSAPKAPAPASALTLPSPVAPLP 1920
QY 1048 EVNPKSDADAGAGETPHMLLRPHVFMPEVTPDMDVLPV-----PNQGIIN-- 1095
DB 1921 PKQPLESAPGSLVLEFSKL-----PVPAEDELPLIPPEAVSGGFEFFQFILVNP 1972
QY 1096 -----PML 1098
DB 1973 APKPAAGTAPAPSAKQPV 1991


```

Db      7436 VPEVKEVVEKKAAPVKKPEVPAKVEPKVLEKPAVPVPERAESPPPEVVEPE 7495
QY      834 ESTPE-----LSAETPKALENSKPEG-----VPTTKPAATKPEMTTAK--- 875
Db      7496 ELAPEEETAPBEKEVPVVAEEBEPVPPPAVPEPKKIPEKKVVPVKKPAAPPKPEP 7555
QY      876 DKITER-DLRTTPTTTAAP-----KMTKETATTTEKTTESKITATTT 917
Db      7556 EKVIEKPKLKRPPPPPPAPPKEDVKEKIFOLKAIPKPKVPENQVPEKV----- 7605
QY      918 QVTSITTQDTTFFKIT--TLKTTTLAP-KVTTTKTITTTIMNKPBTAKPK-----D 968
Db      7606 -----ELTFLKVPGEKKVKRLKPKPEKPEEVLKSVLRKPEEBEPKVEPKKLE 7657
QY      969 RATNSKATPKPKPTKPKPTSTKPKKTPRVRKPKTTTTPRKMTSTMPELNPTGRI- 1027
Db      7658 KVKKPAVPEPPPKPEVEVEVPTVTKREKIPETK-----VPEIKPAIPLP 7704
QY      1028 -----AEAMLQ-----TTRPNQTPNSKLVVEV-----NPKSEDA----- 1056
Db      7705 APEPKPEAEVKTITKPPPEPEPTPIAAVTPVVGKKAEEKAPKEEAAKPKGPIKGV 7764
QY      1057 -----GGAEGTTPHMLLRPHVFMPEVTPDMYDLPRVP 1088
Db      7765 KKTSPPIEAERKLRPGSGGKPE-----PDEAPFTYQLKAVP 7801

RESULT 7
T45025
mucin MUC5B, tracheobronchial [imported] - human (fragment)
C.Species: Homo sapiens (man)
C.Date: 21-Jan-2000 #sequence_revision 21-Jan-2000 #text_change 21-Jul-2000
C.Accession: T45025
R.Desseyn, J.L.; Guyonnet-Duperat, V.; Porchet, N.; Aubert, J.P.; Laine, A.
J. Biol. Chem. 272, 3168-3178, 1997
A.Title: Human mucin gene MUC5B, the 10.7 kb large central exon encodes various alternat
A.Reference number: Z22899; MUID:97166151; PMID:9013550
A.Accession: T45025
A.Status: preliminary; translated from GB/EMBL/DBJ
A.Molecule type: DNA
A.Residues: 1-3570 <DES>
A.Cross-references: EMBL:Z72496; NID:g1834502; PIDN:CAA96577.1; PID:g1834503
A.Experimental source: placenta
C.Genetics:
A.Gene: MUC5B

Query Match      17.0%; Score 994.8; DB 2; Length 3570;
Best Local Similarity 13.5%; Pred. No. 4e-12;
Matches 445; Conservative 115; Mismatches 459; Indels 2282; Gaps 98;

QY      6 LPYLLLLLVFVIOQVSSQLSCK-----GRCFESPER----- 39
Db      3 LPV-----STVCREV-----CAWSWYNGHRPEGLGGDFETPENLRQRYQVCPV 50
QY      40 -----GRECDCAQ-----CKYDKCCP 57
Db      51 LADIECRAAQLPDMLEBLGQQVDCDRMRGLMCANSQQSPPLCHDYELRVLCCEYVPCGP 110
QY      58 ----- 57
Db      11 SPAGTSPQPSLASTEFAVPTPTQTATETKTLWVTPSINSTAALTSQGGSSGPGVTVT 170
QY      58 -----DY-----ESF 62
Db      171 PSAPGTTTCQPCRWTEWFDSDYPKSEQLGQGVESYDKIRAAGGHLCOQPKDIBCOAESF 230
QY      63 -----CAEVH----- 67
Db      231 PNWTLAQGVQKHCDVHFLGVCNRWEGVFMKYNVIRVLCSSDDHCHGRATTPPPPT 290
QY      68 -----NPTSPSSKAPPS-----GA 84

```

```

Db      291 ELETATTTTQALFSTPOFTSPTSGLTTRAPPASTTAVPTLSEGLTSPTSTLTGTTATGGP 350
QY      85 SOTIKST-----TKRSPKPPNKKTKKKVIE----- 109
Db      351 ROSAGSTBPTVGVATSLTPRSALPGTTSGLTWRPQPPPLATPTMATSRARPTGTAS 410
QY      110 ---SEE-----ITEHVSVENQES----- 125
Db      411 TASKPLATTSLATTLTSELSTSQABTSTPRTBTWSPLTNTTSQGTTRCQPKCEWTF 470
QY      126 ----- 125
Db      471 DVDFTSGVASGDMETFFENIRAAAGKMCWAPKSI ECR AENYPEVSI DQGVQLTCSLTG 530
QY      126 -----SSSSSSSSSSSTTIWIKISSKNSA 148
Db      531 LTCKNEQDQGRNMCNVRVLCDDYSHCPSTLTATSTATPSSTPGTTMIL----- 583
QY      149 ANRELQKLVKDKNKKRTKPKPPVDEAGSLDNGDFKVT--TPDTSTTQHNKVS 206
Db      584 -----TKPTTATTAGSTATASSTQATAGPHVSTT-----A 618
QY      207 TSPKIITAK-----PINRPSLPNPSDTSKETS----- 235
Db      619 TTPVTSSKATPPSPGATAPALRSTATTPTATSTFTAI PSSLGTTWRLSQTTTMA 678
QY      236 -----TVNKEITVETKETTT-----TNKQSTDGKEKTTSAKETQSI EKTSAK 278
Db      679 TMTATSPSTPETVHTSTVLTATTATTGATGVSATPSSTPGTAHTTKVLTITTTGFTATP 738
QY      279 DLAP-----TSKVL-----AKETPKAE 295
Db      739 SSSPGEARLPVWISITTTTPTTRGTVTPSSIPGTHPTVLTITTTTAVATGSMATPSS 798
QY      296 TTTKG--PALTTPKPEPTPTT-----PKPEASTTPKEPTPTTIKAP----- 334
Db      799 TOTSGTPPSLTT-----TATTITATGSTNPSSTPGTTPPIPPVLTATTATPAATSTVTPS 854
QY      335 ---TTPKEPAPTTKS-----APTTP-----KEPAPTTKKPA 364
Db      855 SALGTHITPPVENTTATTHGRSLSPSSPHTVCTANTSATSGILGTHITEPSTGSHTPA 914
QY      365 PTT-----PKSPAPTTPKPEAPTTPKSAPTT-----PK 392
Db      915 AITGTTQHSTPALSPSPSSRTTESPPSGITTPGHTTATGRTATATATPSKTR--STLLPS 974
QY      393 EP--APT----- 398
Db      975 QPSTAPITVTVMGCEPQCMSEWLDYVPMGPGSGGDFDYSNIRAAGGAVCEQPLGLE 1034
QY      399 -----PKXP-----APTTPK 408
Db      1035 CRAAQPGVPLRELQGVCECLDFGLVCRNREQVKFCMCFNYEIRVFCNKGHCFSTPA 1094
QY      409 EPAPTTPKE-----PTPTTPKEPAP-----TTKEAPT----- 436
Db      1095 TSTATSPSTPGTWTILTELITTTATTTESTGSTATPTSTLRAPPKVLTTATTPTVTS 1154
QY      437 ---TPKEAPTAPKKA-----PTTPK-----EPAP-----TTKEAPTPTTK 471
Db      1155 SKATPSSSGTATAPALRSTATTATATSVIPI PSSLGTTWRLSQTTTPTATMSTATP 1214
QY      472 BPSP-----TTPKEAPTTPKSAPTTPKPEAPT-----KSAPTTP 508
Db      1215 SSTPETAHTSTVLTATATTGATGVSATPSSTPGTAHTTKVPTTTTGTATPSSSGTA 1274
QY      509 KSP-----SPTT----- 515
Db      1275 LTFPPVWISITTTTPTTRGTVTPSSIPGTHATVLTITTTTAVATGSMATPSSSTQTSCTP 1334
QY      516 -----TKSPAPTTPKPEAP-----TTPKKAPPT-----TPKEAPTTPKPEA 552
Db      1335 PSLTTTATTATGTTNPSSTPGTRPPIPPVLTATTATPAATSTSTVTPSSALGTHTPPV 1394

```


A:Title: Gene fusion is a possible mechanism underlying the evolution of STA1.

A:Reference number: A91831; MUID:87194600; PMID:3106330

A:Accession: A26877

A:Molecule type: DNA

A:Residues: 1-242 <YAM>

A:Cross-references: EMBL:M16164; NID:g172522; PIDN:AAA35014.1; PID:g172525

A:Accession: B26877

A:Molecule type: DNA

A:Residues: 762-1331 <YA2>

A:Cross-references: EMBL:M16165; NID:g172523; PIDN:AAA35015.1; PID:g172526

R:Parado, J.M.; Ianez, E.; Zalacain, M.; Claros, M.G.; Jimenez, A.

FBS Lett. 239, 179-184, 1988

A:Title: Similar short elements in the 5' regions of the STA2 and SGA genes from *Saccharomyces cerevisiae*

A:Reference number: S27281; MUID:89031230; PMID:3141213

A:Accession: S27281

A:Molecule type: DNA

A:Residues: 1-31 <PAR>

A:Cross-references: EMBL:X13857; NID:g4551; PIDN:CAA32069.1; PID:g4552

R:Lambrechts, M.G.; Bauer, F.F.; Marmur, J.; Pretorius, I.S.

Proc. Natl. Acad. Sci. U.S.A. 93, 8419-8424, 1996

A:Title: Muc1, a mucin-like protein that is regulated by Mss10, is critical for pseudohyphal growth in *Saccharomyces cerevisiae*

A:Reference number: JG6123; MUID:96323217; PMID:9710886

A:Accession: JG6123

A>Status: nucleic acid sequence not shown

A:Molecule type: DNA

A:Residues: 1-1367 <LAM>

A:Cross-references: GB:U30626; NID:g1304386; PIDN:AAC49609.1; PID:g1304387

C:Genetics:

A:Gene: SGD:MUC1; STA2; MAL5; DEX2; SGD:S0001458

A:Cross-references: MIPS:Y1R019c; SGD:S0001458

A:Map position: 9R

C:Superfamily: yeast glucan 1,4-alpha-glucosidase homolog; glucan 1,4-alpha-glucosidase

C:Keywords: glucanase; hydrolase; polysaccharide degradation; transmembrane protein

F:5-21/Domain: transmembrane #status predicted <TM>

F:1350-1366/Domain: transmembrane #status predicted <TM2>

Query Match 16.0%; Score 933.8; DB 1; Length 1367;

Best Local Similarity 25.6%; Pred. No. 1.6e-11;

Matches 350; Conservative 110; Mismatches 446; Indels 462; Gaps 69;

QY 9 YLL-----LLLSVFVQVSSQBLSCGRC-----CQWMPNFQCFEVLQSSAAQYASSQWG 169

DB 118 YLLDNPDTDFATFVATQDVS-----QVWMPNFQCFEVLQSSAAQYASSQWG 169

QY 36 --SFERGECDDAOKKYDKCCPDYEFCAEVHNPSPSSKAPPSSGASQT----- 87

DB 170 TTSF-----DLSTGCNNYN-----QHSQTDFFGFY 196

QY 88 -----IKSTKSPKPPNKKTKKVVSEBEITEHSVSENQSSSSSSSSSSSS 136

DB 197 WNIIDCNNCGTKSSITSSSESTSSSTSSSTSSSTSSSTSSSTSSSTSSSTSSST 256

QY 137 TIWKIKSSKNSAANRELQKKLVKDNKKNRKKKPT-----PKPPVDEAGSGL 185

DB 257 TAPATPTTSC-----TKRKPTPTTSCCKKPTPPHH----- 291

QY 186 DNGDFKVTPTDSTTQHNKVSFSPKITTAKPINRP-----SLP-----PNSDTSKE 232

DB 292 -----TTPCT-----KKTTTSKTCYKKTTPVPTSSSTSSSAPVTPSSSTTES 339

QY 233 TSLVYNKETT-----VETKETTNNKQTS-----TDGKETTSAKETQSIKTSAXDLAPTS 284

DB 340 SSAPVTSSTSSSAPVTPSSSTSSSAPVTSSTSSSAPVTSSTSSSAPVTSSTSSS 395

QY 285 KVLAKPTKAEHTTKGPA-----LTPKEPTTPKPASTTPKPTPT-----IKGAP- 334

DB 396 -----PTSSSTSSSAPVTSSTSSSAPV-----SSTSSSAPVTSSTSSSAPV 446

QY 335 ---TTPKEPAPTTKSAPTTKKPAPTTKKPAPTTKKPAPTTKKPAPTTKKPAPTT 391

DB 447 TSSTTESSSAPVTPSSSTSSSAPVTS-----TTESSSAPVTPSSSTSSSAPV 501

QY 392 KEPAPTTKAP-----TTPKEPAPTTKAPT-----TTPKEPAPTTKAPTTPKEP 441

502 ---SSTTESSAPVTPSSSTSSSAPV-----PTPSSSTTESSSAPV-----SSTTESS 551

QY 442 APTAPKKPAPTTKKEPAPTT-----KAPAPTTTKEPSTTPKKEPAPTTTSAPTT 492

DB 552 APV-----PTPSSSTTESSSTPTSSSTSSSAPVTPS-----SSTTESSAPVTPSSSTT 604

QY 493 TKEPAPTTTSAPTTKKEPSTTPKKEPAPTTKKEPAPT-TPKKPAPTTKKEPAPTTKKEP 551

DB 605 ESSSAPAPTPSSSTTESSSAPVTS-----TTESSSAPVTPSSSTTESSSAPVTP-----P 656

QY 552 APTTTKAPAPTTKKEPAPTTKKEPAPTTPKKTUPTTPEKLAPTTPEKLAPTTPEKLAPTT 611

DB 657 SSSTTESSSAPVTPSSSTTESSSAPV-----SSTTESSSAPV-----SSTTESSSAPV- 707

QY 612 PEPSTP-----TTPKEPAPT-TPKAAAPNTKKEPAP-----TTPKEPAPTTPKKEPAPTTPKE 662

DB 708 ---PTPSSSTTESSSAPVTPSSSTTESSSAPVTPSSSTTESSSAPV-----SSTTESS 760

QY 663 TAPT-TPKGTAPTTLKKEPAPTTKKEPAPKELAPTTTKEPTSTSDK---PAPTTPKGTAP 718

DB 761 SAPVTPSSSTTESSSAPVTPSSSTTESSSAPVPT--PSSSTTESSSAPVTP--TPSSSSN 817

QY 719 TTPKEPAPT-----TPKEPAPTTPKGTAP-----TTLKKEPAPTTKKEPAPKELA 762

DB 818 ITSSAPSTTPSSSTTESSSAPVTPSSSTTESSSAPVSSSTTESSSAPV-----PTPSSSS 873

QY 763 PTTTKGP-----TSSTSDKAPAPTTPKKEPAPTTKKEPAPTTKKEPAPTTPET--PPTTSE 815

DB 874 NITSSAPSSIPFSSSTTESSTGTT---VTPSSSKYPSQSTETSVSSTTTTIVPKTITS 930

QY 816 VSTPTTKAPTTHKSPDSESTPELAGAPTPKALENSKPKGVPVTKTAPATKEMTT--- 872

DB 931 VTPSTTTTITVTCSTGTNSAGETTSKSPKVTIT---VPTTTTTSVTSSTTTT 985

QY 873 ---TAKDXTTBERDLRTTPT--TAAAPKMTKATTTTETKTESKITATTQVTS--TTO 925

DB 986 TVCSTGTNSAGETTSKSPKTTTTPVCSSTSESTASESTTSPPTPTVTVVSTVITE 1045

QY 926 DTPPFK-----ITLTKTTTLA--PKVTT---TKKTTTTEI---MNKPEET 963

DB 1046 YSTSTKPGGEITTTFTVKNIPTTYLTIAPTESVITVTNFTPTTITTTVCSTGTNSAGET 1105

QY 964 ---AKPK-----DRATNS--KATN----- 977

DB 1106 TSCSPKTTVTVPCSTGTGEYTTTATLTVTAVTTVTVTSSSTGTNSAGTTTGYTK 1165

QY 978 --PKPKPTKAPKKP-----TSTKPKPTM----- 999

DB 1166 SVPTTVVTVLAPSAVTPATNAVPTTITTECSAATNAAGETTSVCSAKTIVSSASAGEN 1225

QY 1000 ---PRVAKPTTTPKPKWTSTMPELN-----PTSRIEAM----- 1031

DB 1226 TAPSATPTVTAIPTTITVITTESVGTNSAGETTTGYTKSIFPTTITLIPGNSGAKNYE 1285

QY 1032 -LQTTTRPQNTENSKLVEVNPKESEDAGGAEGETPHMLLRPHVFMPEVT 1078

DB 1286 TVATATNP-----ISIKTTSQATLATTASAS-----VAPVTVTSPLT 1321

RESULT 9

T34513

hypothetical protein ZK783.1 - *Caenorhabditis elegans*

C:Species: *Caenorhabditis elegans*

C:Date: 29-Oct-1999 #sequence_revision 29-Oct-1999 #text_change 09-Jul-2004

C:Accession: T34513

R:Favell, A.; Vaudin, M.

submitted to the EMBL Data Library, August 1994

A:Description: The sequence of *C. elegans* cosmid ZK783.

A:Reference number: Z21536

A:Accession: T34513

A>Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-3507 <FAV>
A:Cross-references: UNIPROT:Q23587; EMBL:U13646; PIDN:AAC24418.1; GSPDB:GN00021; CESP:ZK
A:Experimental source: strain Bristol N2; clone ZK783
C:Genetics:
A:Gene: CESP-ZK783.1
A:Map position: 3
A:Introns: 14/1; 48/2; 84/1; 196/3; 303/1; 381/1; 586/1; 605/1; 1175/3; 1207/1; 1409/2;
3504/1

Query Match 15.4%; Score 896.2; DB 2; Length 3507;
Best Local Similarity 14.9%; Pred. No. 3.4e-10;
Matches 386; Conservative 177; Mismatches 444; Indels 1578; Gaps 84;

QY 19 IQQVSSQELSCKGRFESPE-RGRBCDDA-----CKKYDKC----- 55
DB 506 VELTTSGLACTSYCPNPSECVGYCECVSGYGNALVGCEDIDCBTEICNIEANECV 565
QY 56 -----CPDY----- 59
DB 566 NLIIGFVCCNPNTAHDDCIDFLTKVIYAYMIIFLLKGLIETKEGLHVLVIGNEEDTV 625
QY 60 -----ESFCAEVH----- 67
DB 626 VATRSNHSSTDLITQVQSRNFSSTGQIILTRGVSGEAVTQTDADEFLGLEISAADL 685
QY 68 -----NPTSPSSKKAPPPSGA-----SOTIKST-- 91
DB 686 AGSGGCIPTLPLEPKIEGSKKA---SGVWTEDEGEDEDLMEEGSGSWTTINGTGI 742
QY 92 -----TKSPKPPKKKTKKVIES---EEITEHSVSEN 122
DB 743 TGSPRSEGINRVITLGLGDGEPEATKPGISAPDKTGEKSXESDGEKLVKDGKEA 802
QY 123 QESSSSSSSS----- 132
DB 803 QSSGSSATSGRKKSEATSGSSSSSAKSGTGSEASGSSGASSSSGSGVSGSSVSTESG 862
QY 133 -----SSSTTIWKIKSKN----- 146
DB 863 SFGTSSSGVSGEATSGTVDGSESGKPKSKSTEEKLPFKNGEKSPISGSDTTGKESSE 922
QY 147 -----SAANRELQKLV----- 159
DB 923 ETTSRKPIEGSDSLTEGSGGEWFETGSKGHFESGKSVSTGKGTQSGAEGSGGPKV 982
QY 160 -----KDNK----- 163
DB 983 PKGFGAPEITTDGEBSSTSTGDKSGKPKADKSDKNVPKTGDKNPDIITDGEDSTSETS 1042
QY 164 -----KNRTK-----KKTTP-----KPPVVDGAGSLDN 187
DB 1043 GGGQGGKSKSGOPPGDKGSEVKKPTSEVDGPNLGGTGGKSNVPLKPTDLP EEGSGI-- 1100
QY 188 GDFKVTTPD---TSTTQH-----NKVSTSPKITTAKPINRPSLSPNDSSTKE 232
DB 1101 -----LTSSGGKNSTFEHGKLERLPKPKTEDKSSETPQLGLEISAGKPE--PEDGTSKB 1154
QY 233 TSLTVNKETVETKETTTNKOT-----STDGKKTTSAKET 269
DB 1155 VGLEIILWEST--TPGSTLLSDSVGLEISGSLTKATKPHVEIEGSGTGDEIAT--T 1210
QY 270 QSIKXTSAK-----DLAPTSKVLAKPTPKAETTKGPALTTPKEPT----- 310
DB 1211 RDVSKSTKPRVVDGDNGETSGVDGKPTTAPTPSSAESSSTGRIPPTSASPEGSGG 1270
QY 311 -----PITPK---EPASTTKPEPTTIKSAPTTPKEPATTKSAPTTP---KEPATTP 359
DB 1271 EAGVPSPDGSGESSSTAFDGVSGPTSSATAPEVP-----TTSASSTPDVAEEGIGPT 1323
QY 360 TK---BPAPTPKPEAPTT-----TKEPAPTTKSAPTTPKPEAPTTTP 399
DB 1324 SKPTASPLETT---APSTEVTSPEGSGTEESTLPPTEGSGSTTSSAPTIV--EPATVLP 1377

QY 400 -----KKPAPTTPKPEAPTT----- 414
DB 1378 PQNRNEXPEPTKDTFALTPTTTTGAPQANDSSVENTKCTSSDECCLDALCERRTGCRCPE 1437
QY 415 -----PKE-----PT----- 419
DB 1438 GFEGAPPKSCVDVDECATGDHNCHEARSQNVYGVYACFCPTGFRKADGSCODIDECT 1497
QY 420 -----PTTPKEPAPTTK----- 431
DB 1498 EHNSTCCGANAKCNKPKGTYSCECENGFLGDGYQCVPTT-KKPCDSTQSSKSHCESNMS 1556
QY 432 -----EPAP----- 435
DB 1557 CEVTVDSVECKEMGKYGKSGKVCEDINECVAEKAPCSLNANCNVNMNNTFFSCSKQGY 1616
QY 436 -----TTPKE----- 440
DB 1617 RGDGFMCTDINECDEHHPCHPABCTNLEGSFKCECHSGFEGDGIKKCTNPLERSCEDVE 1676
QY 441 -----PA----- 442
DB 1677 KFCGRVDHVSCLSVRIYNGSLSSVCECEPGRFPEKESNCVDIDECESRNCCDPASAVC 1736
QY 443 ----- 442
DB 1737 VNTGSRCECAEYEGEGVCTDIDECRGMAGCDSMAMCINRMGSGCGCKMAGYTGDG 1796
QY 443 -----KPAPTTPKPEAP----- 459
DB 1797 ATCIKIBEEPKSKDTACTDEWSRLCELEKKQCTVDEBEVPQCGACLPGHHPINGTCOSLQ 1856
QY 460 ----- 459
DB 1857 ISGLCAQKNCNKAEACIDIHDSHFCSGCDPGFIGDGMICDDVDCECNAGMCDDENTKCE 1916
QY 460 -----TTPKE-----APTTPKEPPTTPKEPAPT 484
DB 1917 NTIGSFNCVLEGGFKYDKCVDEKKQPNREKIEIDEENSSSSSSSOEKPTTKGIVSST 1976
QY 485 TTKSAPTTTKEPAPTTTTSKAPTTTKEPPTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAP 544
DB 1977 SATSESTTAEPHVTTISSTTSTKDMTSSKSPENVTMSSESPEVSTSSSKSTTASSETTV 2036
QY 545 PTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTT 595
DB 2037 SSTSESS-----SSEAPLTSSPATTEV--ITESSVKSTTPKESSSSEITVKLSKS 2087
QY 596 PE-----KPAPTTPEELAPTTPEEPPTTPEEPAPTTTPEEPAPTTTPEEPAPTTTPEEP 643
DB 2088 PEVTSSVKSSPSTPS-----TTSQSVSTVPETSKSTVLSSEAPVSTSTSEVHTSSETK 2143
QY 644 PKEPAPTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAP 693
DB 2144 PSLSASITGDTNSTTSTSSLASVKSAPETSASVAPVKLSLSPDV--SQSTKTFFD 2202
QY 694 PT-----TTKEP-----TSTTSDKPAETTPKGATTPKEPAPTTTKEPAPTTTKEPAP 730
DB 2203 ATESSTVQASSTSGTSVKSTSEDESHVTKLSITSSNPSSSVPTSPKSTPTVPEST--E 2260
QY 731 EPAPTTPKG-----TAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTT 756
DB 2261 QFTTTSQGLSTPMNNSSEVLTTSEPHVLSLSPDVSSSTTNNLSSESTVETPKTS 2320
QY 757 APKEL---APTTPKGT-----STT-----SDKP----- 777
DB 2321 SEVLSNKEEPTTTEAPTTLSPDILSTTNNLSQSVSTVSTEDRSRISSENSEKPTSAPELV 2380
QY 778 -----APTTP-----KETAPTT-----KEPAPTT 797
DB 2381 TSSVTHVASSPDVPTESSEPDLLTGSSTENIPEASSKQTSISSTPTDPTTASSEPTKST 2440
QY 798 PKKP-----APTTP----- 807

| | | |
|---|---|-----|
| T29757 | protein UNC-89 - Caenorhabditis elegans | 284 |
| C:Species: | Caenorhabditis elegans | |
| C:Date: | 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 03-Dec-1999 | |
| C:Accession: | T29757 | |
| R:Du, Z.; Le, J.T.; Wilson, R. | | |
| submitted to the EMBL Data Library, May 1997 | | |
| A:Description: | The sequence of C. elegans cosmid C09D1. | |
| A:Reference number: | Z20679 | |
| A:Accession: | T29757 | |
| A:Status: | preliminary; translated from GB/EMBL/DBJ | |
| A:Molecule type: | DNA | |
| A:Residues: | 1-6642 <DUZ> | |
| A:Cross-references: | EMBL:AF003131; PIDN:AAE54132.1; GSPDB:GN00019; CBSP:unc-89 | |
| A:Experimental source: | strain Bristol N2; clone C09D1 | |
| C:Genetics: | | |
| A:Gene: | CBSP:unc-89 | |
| A:Map position: | 1 | |
| A:Introns: | 17/2; 108/3; 154/2; 211/2; 265/3; 326/2; 352/3; 426/2; 454/1; 500/1; 537/1; 6 | |
| /3; 591/1; 6027/1; 6061/3; 6153/2; 6515/1; 6552/3; 6609/1 | | |
| Query Match | 15.1%; Score 879.9; DB 2; Length 6642; | |
| Best Local Similarity | 17.1%; Pred. No. 1.8e-09; | |
| Matches | 379; Conservative 173; Mismatches 419; Indels 1251; Gaps 89; | |
| QY | 9 YLLLLSVFVIOQVSSQE-----LSCKGRCFESFERGREGC---DCDAQCKK 51 | |
| DB | 393 YSIRLDKYNIRHHTTDETVLQPOEGLPSFRPKDFETSEYVRKAWLEDIAEEQEK 452 | |
| QY | 52 Y-----DKCCPDYEFSCARVHNPT-----SPSSK-K 77 | |
| DB | 453 YAAERDAISMTATSEMTASSVDFMNASDQSEFSEWSGSRKSLFFPGPEGGPRKKVK 512 | |
| QY | 78 APP---PSGASOTIKS-----TTKRS-----95 | |
| DB | 513 SPVVISPTGSSISVYGSSSIDWTTGTTLEMQGTRVTRTOYGRFTLQESSAKNCLKVT 572 | |
| QY | 96 --PKPPNKKTKKVBSEBIT-----EHSVSENQBS-----125 | |
| DB | 573 GYPLP-----DITWKDDVQLHEDERTFYSDGEGFTAMTIDPVQVDTGKY 619 | |
| QY | 126 ---SSSSSSSSSTTWIKSSKNAA-----NRELQK-----155 | |
| DB | 620 TCMATNEYQOASTAFPRVLKVEKEAAPAFVTKURDKCEKGDVIDFECEVGVWPEPEL 679 | |
| QY | 156 -----KLKVKDNK-----163 | |
| DB | 680 VMLVDDQPLRSHDFLQYDGTAKLEIRDAQDDTGVTVVKIQNEFGSIEKAEFLVQA 739 | |
| QY | 164 --KN-----165 | |
| DB | 740 DEDKNHVAPEFOATIEYVECEGEYRFRKSVITGDPNPEIIFWFKPLSESEKVFISE 799 | |
| QY | 166 -----RTKKKPTPKPPVVD-----180 | |
| DB | 800 DGICILTIKDVTHFDGWTGQSNLGSASCDGRKLVAVPAPPTFNKPLSDKTVQEK 859 | |
| QY | 181 -----AGSLDNGD-----FKVTTDPDSTTQHN-----203 | |
| DB | 860 TVVFEVDVSGWPPTLTFTLCGKELXNGEGVEIVGHDFYRISIPNTSMXHDGRIYAK 919 | |
| QY | 204 -----KVSTSPKLTAKPI-----NPRPSL 223 | |
| DB | 920 AQNEHGTASRLTVQEEESRSAP--TFLKDIEDQVTKGFAVFETTVRGNPNDEV 977 | |
| QY | 224 -----223 | |
| DB | 978 TWFINGHKMDQSGPGVKIEAHNHDHKLITIDSAQYAGTVLCRAENAVGRFETKARLVVLAP 1037 | |
| QY | 224 -----PPNSDTSKETSILTNNKETT-----ETKEITT-----250 | |
| DB | 1038 EKQKKPP-----KFVEILVDKTEITVDNTVVFEVRVEGEKPTVTWYLKGEELKQSDRVEI 1092 | |

| | | |
|----|---|--|
| QY | 251 -----TNKOTSDGKEKTTSAKETOSIEKTSAKDLAPTS 284 | |
| DB | 1093 REFDSIKISIKIKIEDAGEIRAVATNSEGSDTKAKLTQVKRPAPF-----FDLRPVS 1148 | |
| QY | 285 KVLAK-----PTP-----KAETTKGPAULT-----305 | |
| DB | 1149 LTVEKGEAVFSAHAFGLPLPTVEMVNGRKVRDGOEGARVTRDESDVDCASILTIDTAT 1208 | |
| QY | 306 -----PKBPTPTTPKEPAST-----T 321 | |
| DB | 1209 YYSVNNHLTISVAENLTGABETGAQLTIEPKKESVVVEKODLSSSVQKEIAQOVKEAS 1268 | |
| QY | 322 PKEPTPTTIKSAPTTTPKEPAPTTT-----KSAPT-----350 | |
| DB | 1269 PEATTTITMETSLSSTXTTMTSITETVSTYGGVTVETKESESESATVIGGGSGGVTEGS 1328 | |
| QY | 351 -----TPKEPAPTTTPKEPAP-----TTPK3-----PAPTTTPKEPAPT 382 | |
| DB | 1329 ISVSKIEWSKTSDQTVREGTRRRVSAFEEELPKEVIDSDRKKKXSPSP--DKKEKSPE 1387 | |
| QY | 383 TTKSAPTTTPKEPAPTTTPKAPPTTPKEPAPTTTPKAPPTTPKAPPTTPKAPPTTP--KE 440 | |
| DB | 1388 KTEKPASPTKKTGEEVKSP---KEKSPASPTKKEKSPAAEEVKSPTKKESPSPTKKE 1444 | |
| QY | 441 PAPTAPKAPAPTTTPKEPAPTTTPKEPAPTTTPKESPTTP-----KEPAP-----483 | |
| DB | 1445 KSPSPPTKKTGDEVKEKSP--PKSP---TKKESPEKPEDVKSPVKKEKSPDATNIVEVS 1499 | |
| QY | 484 -----TKAPAPTTTKSAPTTP-----KEPSPPTTPKEPAP 521 | |
| DB | 1500 SETTIEKTTMTTMTHESEERTSVKKEKTEKVEDEKPKSPTKKDKSEKSTIEIKS 1559 | |
| QY | 522 TTPKEPAP--TTPKAPAPTTTPKEPAPTTTPKAPTTTKK-----PAPT-----562 | |
| DB | 1560 PVKKEKSPKVEEPASPTKKEKSPKAPSP---TKKSENVKSPTKKESPEKSVVEE 1615 | |
| QY | 563 --APKEPAPTTPEKATPTPKLTPPTPEKLAPTTPKPA-----PTTEELAPTTPEE 614 | |
| DB | 1616 LKSPKESKSP---EKADDKPK--SPTKKEK---SPEKATEDVKSPTKKESPEKVEEK 1665 | |
| QY | 615 PTPPTPEPAPT--TPKAAAPNTPEKAPPTTPKE--PAPTTPKEPAP-----TTPKET 663 | |
| DB | 1666 PTFKESKESPTKKTDDVKSPTKKESKQTVBEKSPASPTKKEKSPKSVVVEKSPK 1725 | |
| QY | 664 APPTPKGTAPTTLKEPAPTTTPKAPKAPKELAPTTT-----KEPTSTSDKAPPT 711 | |
| DB | 1726 SP-----EKAEEKPKSPTKKEKSPKSAABEVKSPTKKESPEKSAEEKKSP 1773 | |
| QY | 712 TPKGTAPT-----TPKEPAPTTTPKEPAPTTTPKGTAPTTLKEPAPTTTPKAPKAPKEL 761 | |
| DB | 1774 TKKSSPVKXVADDEVKSPTKKESPEKVEE--KPASP-----TKK3---KTPEKSAABEL 1823 | |
| QY | 762 APPTTK-----GPTSTTSD-----KPAPTTTPKETAPTTTPKEPAPTTTPKAPPT 804 | |
| DB | 1824 KSPTKKESKSPSPKTKTGDESKESPEKPEKPKSPTPKKSPPGSPKPKKSKSPKSAEKPP 1863 | |
| QY | 805 TPEPPTTSEVSTPT-----TTKEPTTIHKSPD---833 | |
| DB | 1864 APK-----LTRDLKLQVNTKDLAHFEVVVHATECKWFLDCKEITTAQGVTVSKDDQFEF 1939 | |
| QY | 834 -----ESTPELSAETPKALENSPKPEGVPTTKTFAATKP 868 | |
| DB | 1940 RCSIPTTMFGSGTVSVVNAAGSVETKELKLETFK-----ETK-----KP 1982 | |
| QY | 869 EMTTAKD-----KTTED-----882 | |
| DB | 1983 EFTDKLRDMVEVTKGDTVQMDVIALHSPLYKWNQNLLEDKNGKGVTKNEENKSSLIIPN 2042 | |
| QY | 883 -----LRTTPTETTTA---APXMT---KETATTTEKTTES---910 | |
| DB | 2043 AQDSKGTVEASNEVSGSSSAQLTVNPPSTPIVVDGPKSVTIKETETAEFKATISGFP 2102 | |
| QY | 911 -----KITATTQVTSVTTTQDTPPKIT-----TLKTT-----936 | |

```

Db      2103 APTVKATINEKIVBESRTITTTIKTEDVYTLKISNAKIBQGTGVKVTQAQNSAQDSQADL 2162
QY      939 -----TLAPKVTITTKTIT-----TTE 955
Db      2163 KVEPNVKAPE-----KSLQTDKVADEGEPLRWNLDELDPSPGTEVSWLLNGQPLTKSDIVQ 2219
QY      956 IMN-----KPEE-----TAKPKDRA-----TNSKATPKPQKPTKAPKPTS-- 992
Db      2220 VDHGSGTYHTVIAEAKPEMSGTLTAKAKAAGACEBTSKAVTVNGNKKPEFVQAPQNH 2279
QY      993 -----TKKP-----KTMPEVRKCP-----1005
Db      2280 TTIIEESVKFSAIVTGKPMNVWYLNKKLIOSEBVKVYVHETGKTSIRIQKPLMEHNG 2339
QY      1006 -----KTTPTPRKMSTM-----PE-- 1020
Db      2340 TIRVEAENVGKVQATAQAKVDKTEVP-KFTTNMDRQVKEGEDVKFTANVEGYPEPSV 2398
QY      1021 ---LN--PTSRIAEAMLQTTTRPNQTPNSK---LVEVNPKSEDAGG-----1058
Db      2399 ANTLNGEVPVK-----HPNITVTDKQGEHTIEISAVTPEQAGELSCAETNPFVGSK 2448
QY      1059 -----AEGETPHMLLRPHVFM-----EVTDM 1081
Db      2449 KRQVQLAVKVGDAPTFAKNLEDRLITEGELTMDAKNLNIVKPKKITWLKDGVEITS 2508
QY      1082 DY 1083
Db      2509 HY 2510

RESULT 12
B48666
C:Species: Homo sapiens (man)
C:Date: 03-May-1994 #sequence_revision 03-May-1994 #text_change 15-Mar-2004
C:Accession: B48666
R:Schlueter, C.; Duchrow, M.; Wohlenberg, C.; Becker, M.H.G.; Key, G.; Flad, H.D.; Gerde
J. Cell Biol. 123, 513-522, 1993
A:Title: The cell proliferation-associated antigen of antibody Ki-67: a very large, ubiq
ins.
A:Reference number: A48666; MUID:94043435; PMID:8227122
A:Accession: B48666
A>Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-2897 <SCH>
A:CROSS-references: EMBL:X65551
C:Keywords: alternative splicing; cell cycle control; nucleus; tandem repeat
F:29-91/Domain: kinase interaction domain homology <KIH>

Query Match 15.0%; Score 875.8; DB 2; Length 2897;
Best Local Similarity 14.9%; Pred. No. 6.5e-10;
Matches 427; Conservative 138; Mismatches 439; Indels 1862; Gaps 104;

QY      3 WKT-----LPIYLLLLSVFVIQVSSQELSCRCFESPERGECDCDA 47
Db      2 WPTRLVTIKRSGVDPHFPLSLTCL-----FGGIECDIRI 39
QY      48 Q-----CKYDKCCPDYSEFCAEVH-----NPT-----70
Db      40 QLPVVKQHQCK-----IEIHEQAILHNFSSNTQVNGSVIDEPVRLKHGD 86
QY      71 -----SPPSSKKAPPPSGASQT 87
Db      87 VITIIDRFYENESLONGRKSTFPRKIREQEPARRVSRSSFSDDPEKAQDSKAYSKI 146
QY      88 IKSTTKRSPK-----PPN-----KKKT-----KKVIESEBIEEHSVSEN 122
Db      147 TEGKVSNGNELFDENLPNTPLKRGAPTKEKSLVMHTPPVLKILK-----EQPQPSGK 201
QY      123 QESS-----SSSSSSSSST-----137

```

```

Db      202 QESGEIHVEVKAQSLVISPPAPSPKTPVADQRRRSCKTAPASSSKSQTEVPKRGGER 261
QY      138 -----IWKIKSSNSAANRELQK-----KLKVKDNK-----163
Db      262 VATCLOKRVISRSQHDILQMICSKRRSGASEANLIVAKSWADVVLGAKQQTQTKVIKHG 321
QY      164 -----KNRTKKKP-TPKPPV-----177
Db      322 PORSMMKQRRPATPKPKPGEVHSQFSTGHANSPTIIIGKAHTEKVHVPAFYPVRLNF 381
QY      178 -----VDEAG-----SGLDNGD-----189
Db      382 ISNQKMDFKEDLSGIAEMFKTPVKEQPOLTSTCHIAISNSENLLGKQFQSTGSGEPLJP 441
QY      190 -----PKVT-----TPD 196
Db      442 TSESGFNVFSAQNAAKOPSKCSASPLRQCIENGNGVAKTPRNTYKMTSLETKTS 501
QY      197 TSTTQHNKYST---SPKITTAKEPNRP-----221
Db      502 TETEPSKTVSTVNRSGRSTEFNRNIQKLPIVESKSEETNTEIVEICILKRGOKAILQORREG 561
QY      222 -----SLP-----224
Db      562 EMKEIERPETYKENTELKENDEKMKAMKRSRTWGCKCAPMSDLTDLKSLPDTLMKDTA 621
QY      225 -----PNSDTSKETSLTVNK-----ETTVETK-----246
Db      622 RQNLLOTQDHAKAPKSEKGIKMPQCLOQPEPINTPTHTROQLKASLGKVGVEELIA 681
QY      247 -----ETTTNKQSTDGKE-----KTTGAKETQSIE 273
Db      682 VGKFTSTGETHTHREPAGDGKSIITFKESPKQILDPAARVTGMKWPRTPKBQAQSL 741
QY      274 KTSKDLAPTSKVLAKPTPKAE-----TTTK-----299
Db      742 -----DLAGFKELFQTPGSEESMTDEKTTKIACKSPPEVDPTSTKQWPKRSLRAD 796
QY      300 -----GPALTTPKEP-----TPT-----TPK 315
Db      797 VSEEFIALRLKTPSAGKAMLTTPKAGGDEKDIKAFWGTVPQKLDLAGILPGSKRQLOTPK 856
QY      316 EPAST-----TPKEP-----TPTIKSAFT-----335
Db      857 ERAQALIEDLAGFKELFQTPGHTTEELVAAGKTKIPCDSPQSDPVDPTPTSTKQPKRSIRK 916
QY      336 -----TPKEPAP-----TTTSA 348
Db      917 ADVEGELLACRNLMPSAGKAMHTPK-----PSVGEEDIIIFVGTVPQKLDLENLTGSKRR 973
QY      349 PTPKEPAPT-----TTKEPAPTTPKEPAPTTPKEPAPT 383
Db      974 PTPKEEAQALDITGFKELFQTPGHTTEELVAAGKTKIPCDSPQSDPVDPTPTSTKQPKRSIRK 1027
QY      384 TKSAPTTPKE-----PA 395
Db      1028 TRQPKTPLEKRDVQKELSGALKLTQTSGETHTDKVPGGEDKSIINAFRETAKQLDPAA 1087
QY      396 PTPPKPAPTTPKEPA-----PTTPK-----EPTPT-TPK 424
Db      1088 SVTSGKRHKPT-KERQAQPLEDLGAKWELFQTPVCTDKPTTHEKTTKIACRQSDPVDPT 1146
QY      425 EPAPTKE-----PAPT-----436
Db      1147 SSKPOSKRSLRKVDVEEFFALRKRTPSAGKAMHTPKPAVSGEKNIYAFMGTPVQKLDLT 1206
QY      437 -----TPKEPA-----PTAPKK 448
Db      1207 ENLTGSKRLQTPKEKAQALDLAGFKELFQTRGHTTESMTNDKTAKVACKSSQPDLDKN 1266
QY      449 PA-----PTTPKEPAPT-----460
Db      1267 PASKRRLNTSLGKVGVEKELLAVGLTQTSGETHTHTTEPTGDKSMKAFMSPKQILD 1326

```

```

Qy 461 ---TPKEPA-----PTTKPSPTTPK-----EP-- 481
Db 1327 SAASLTGSKRLRTPKRGSEVPEDLAGPIELFQTSHTKE-SMTNEKTKVYSRASQDPL 1385
Qy 482 --APITTKSAP-----
Db 1386 VDTPTSSKPPKRSRLKADTSEFLAFKQTPSAGKAMHTPKPVGEBKDINTFLGTPVQ 1445
Qy 491 ---TTTKEPA-----PT-----TTK-----SAPTT 507
Db 1446 KLDQGNLPGSNRLQTRKEKAQALEELTGFRELQTPCTDNPTADEKTTKKILCKSPQS 1505
Qy 508 PKBPSPPTTKB-----PA-----
Db 1506 DPADPTTNTKQPKRSKADYVEEFLAFRLKLTSPSAGKAMHTPKAAVGEEDINTFVGT 1565
Qy 521 ---PTTPKEPAPT-----TP-----KPK 535
Db 1566 VEKDLGLNLPGSKRPQTPK3KAKALEDLAGFKELFQTPGHTESMTDDKITEVSCSP 1625
Qy 536 APPTKEPA-----
Db 1626 QPDVPKTPSSKQRLKISLGVGVKEEVLPGKLTQTSKTTQTHRETAGDGKSIKAPKE 1685
Qy 545 ---PTTPKEPA-----
Db 1686 SAKQMLDPANYGTGHERPRTPKKEAQSLDLAGFKELFQTPDHTTESTTDDKTKIACK 1745
Qy 566 EPAP---TTPKET---APTTP-----KKLTPTT----- 587
Db 1746 SPPPSMDPTSTRRPKTLGLKRDIVBELSALKQLTQTHDKVPGDEKGINVRETA 1805
Qy 588 PEKLPAP---TTEKAPAPT-----BELA-----PTPEPTPTTPPEP 623
Db 1806 KQKLDPAASVSGKQRPRTPKGAQPLEDLAGLKFQTPVCTDPTTHEKTKIACRSP 1865
Qy 624 AP---TTPKAAA-----
Db 1866 QPDVPVGTPIFKPQSKRSRLKADVEESLALRKRTPSVGKAMDTPKAGGDEKDKMAPMG 1925
Qy 633 ---PNTPKPEA-----PTTPKEPAPTTPKE 654
Db 1926 TPVQKLDLFGNLPGRKWPQTPKEKAQALEDLAGFKELFQTPGTDKPTTDEKTKIACKS 1985
Qy 655 PAP---TTPKET-----APTTPK----- 669
Db 1986 PQDVPDIPASTKQPKRNLRKADVEEFLALRKRTPSAGKAMDTPKPAVDEKKNITFV 2045
Qy 670 ---GPAPTTLKEPAPTTPKXPAP-----KELAPTTTKEPTSTTSK----- 707
Db 2046 ETPVQKLDLGLNLPKSGKQ--PQTPKEKAQALEDLVGFKELFQTPGHTESMTDDKITEV 2103
Qy 708 ---PAP-----TTPKGTAPT----- 719
Db 2104 SKSPQSFESFTRSRSKQRLKPLVKVDMKEBPLAVSKLRTSGTGTQTHTEPTGDSKSI 2163
Qy 720 ---TPK---BPAPT-----TPKEPA----- 733
Db 2164 KAFKESQKILDPAAVSGSRRELQTRKEKABALEDLVDFKELFAPGHTESMTIDQNT 2223
Qy 734 ---PTTPKGTAPTTLKEPAPTTPKXPAPKELAP-----TTTKGPTSTTSKPA----- 778
Db 2224 KIPCKSPPELDTATSTKRCB-KTRPRKEVKEELSALVERLTQTSQSTHTEKBPASGDE 2282
Qy 779 ---TTPKE-----TAPTT 789
Db 2283 GIKVLQKAKKXPNVEBPSRRRPRAPKEKAQPLEDLAGFTELSETSGHTQESLITAKA 2342
Qy 790 PKEPAPTTPKPAPTTPPTPTTSEVSTPTTKPTTHKSPDESTPELSAETP----- 845
Db 2343 TKIPCESPLEVVDITASTKRLRLTRVQKVQKEBPSAV-KFTQTSGETTADKEPAGED 2401

```

```

Qy 846 ---KALENSKPEFGVPTT-----KTPAA---TKPEMTTAK 875
Db 2402 KGIKALKESAKQTPAPAAVSGSRRRPRAPRESAQALIEDLAGPKDPAAGHTESMT---D 2458
Qy 876 DKITERDLRTPB---TTTAPKMTKETATTTKTE---SKTATTQVTTTQ--- 925
Db 2459 DKITIPCKSPELEDATSSKRPRTAKQVKBELAVGLKLTQTSGETTHTDKBPVG 2518
Qy 926 ---DTPPFKITTLLKTTTLLAPKVVTTTKTITTTIMNKPEETAKPKDRA----- 970
Db 2519 EGKGTAKF-----QPAKENVDAEDVIGSRRRQPRAPKEKAQPLEDLASQEL 2565
Qy 971 ---TNSKATTTKPKQPTK-----APK-KPT----- 991
Db 2566 SQTGHTTEELANGAADSFTSAPKQTPDSGKPLKISRVLPRAPKEVPGDVVSTRDPVKSQ 2625
Qy 992 ---STKKPKTMTPRVRKPKTTP----- 1009
Db 2626 SKSNTSLPPLPFKRGKGKDSVGTGKELRCMPAPEBEIVEELPASKQKVAPRARGKSEP 2685
Qy 1010 T---PRKMTSTMPDELNPTSRIAMLTQTTTRPNO---TPNSKLVEVNPKSEAGGAE 1062
Db 2686 VVIMKRLRTSAKRIEPAEELNSDMKTEEBHKLQDSVPENKGISLRSRQDKTEAQ 2745
Qy 1063 TPHMLL-----REHVFMEPVTDMDYLPVRPNOGIILNP 1096
Db 2746 ITEVFVLAERIEINREKCP---MKTSPMD-----IQNP 2777

RESULT 13
T25697
hypothetical protein F16F9.2 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 09-Jul-2004
C:Accession: T25697
R:Fulton, B.
submitted to the EMBL Data Library, August 1996
A:Description: The sequence of C. elegans cosmid F16F9.
A:Reference number: Z20071
A:Accession: T25697
A>Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-1229 <FUL>
A:Cross-references: UNIPROT:Q94185; EMBL:U67956; FIDN:AA07691.1; GSPDB:GNC0028; CESP:F1
A:Experimental source: strain Bristol N2; clone F16F9
C:Genetics:
A:Gene: CESP:F16F9.2
A:Map position: X
A:Introns: 35/1; 361/1; 384/1; 482/2; 517/1; 971/1; 1021/2; 1179/3

Query Match 14.9%; Score 871.4; DB 2; Length 1229;
Best Local Similarity 25.4%; Pred. No. 2.4e-10;
Matches 346; Conservative 100; Mismatches 360; Indels 556; Gaps 66;

Qy 2 AKWTLPIYLLLSVFIQVSSQELSCG-RCFESFERGRCDCDAQCKKYKCCPDYE 60
Db 4 AW-VVSAFLILGNVVQSSSLSKTINDGSRDFKIVKHKKNTCTCSC---KCVPD-- 56
Qy 61 SFCAEVHNPTSPSSKKAPPPSGASOTIKSTTRSPKPPNKKTKKVIESEETEEHSVS 120
Db 57 ---AFSN-----PFVSVTTISSN-----N 73
Qy 121 ENGESSSSSSSSSSTIW-----KIKSKNSAANRELQKLVKDNKNKRTK 169
Db 74 DNVDIGSGDSNPTGSS-WFOEIEATVGGQTVKSEHNIDSSVEVEKKV----- 120
Qy 170 KPTPKPVVDAGSGLNDGFKVTTTDTSTQ---HNKYSTSPKI----- 211
Db 121 ---TISTDASTNAPTGGKSTTPEITGIVVINSKSESVT 158
Qy 212 ---TTAKPINRPSLPNSDTSKETSFLVNNKETTIVETKETTINKQSTDGKEKT 263
Db 159 DMSITRSTTLSP-----TTELJTSPELTVSTDSSTST-EQTSPDNTTEI 202

```

```
QY 264 TSARETSQIEKTSADKLAPTSKVLAKPTPKAETTTKGPALTTPEPTPTT---PKPEAST 320
Db 203 ASPMET-----NTTTE-ATTTSVEPSVSTLASEDETTVT 235
QY 321 TPKEPTPTTIKSAPTTKEP---APTITKSAPTTKEPAPTITKPEAPTITKPEAPTITK 377
Db 236 AIAESTTTTIVAEVSTTEPTTAESTTKSTT---KAPATTTEPTTTEE---VTTTE 289
QY 378 BPAPTITKSAPTTKEPAPTITP---KKPAPTITKPEAPTITKPEP---TPTTKEPAPTITKE 432
Db 290 AETSITTSSESTSEK---PTTFLDNKIAGATGK---PETTHFPVCTTP-----NFD 337
QY 433 PAPTPP---KEPAPTAPKKPAPTITKPEAPTITKPEAPTITKPE---PSPTTKEPAP 483
Db 338 TATETPFVAKGDKMLSKTAAETTTQOTTEVT-DGPEKETTKVNSIEIPTTVPLVETT 396
QY 484 TTTKSAP-----TTTKEPAPT---TTKSAPTTPKEPSPTTKE---PAFT----- 522
Db 397 STTSTAKESDGFHTLLKLVTTADSDSTESATTVKPFNEETITKSHVVPKTKGTVKV 456
QY 523 TPX-----EPAPTTPKXPAP----- 537
Db 457 TPKLELSPDEPTET-KAPHGKLEKKTIVHFLVLSNDFARYSEAKENDVYNHLDYNYRE 515
QY 538 -----TTPKPBA-----PTTPKPAPTTTKPAPTAPKPA---PTTPKE 574
Db 516 AKEPTTBESESTTEVTTEBPANTGNPTTEN---PTTEQPTSTAESITTTALPFTTEQ 572
QY 575 TAPTPPKLPTTPPKLAPTTPKPAPTTPELAPTTPPEAPTTPPEAPTTPPKAAAPN 634
Db 573 TV-----TTEE---PTTAERSTATQ---KPTIQESVST---EKTSTTKA---S 610
QY 635 TPKEPAPTTPKEPAPTTPKEPAPTTPKPTAP---TPKGTAPTTLKEPAPTTPKPAKPE 691
Db 611 TTEE---PTTTOBPTTT---ESSTTGKATTPELSTSEITTELT---ITE----- 655
QY 692 LAPTTKPTSTSDKPAP---TPKGTAPTTPKEPAPTTPKPAPTTPKGTAPTTLKE 747
Db 656 -GSTTTEPTTAFAEASTGIIITDEETSTSTTBEITSKE---IVTESAITQTSVS 712
QY 748 PAPTPP-----KKPAPKELAPTITKGP----- 769
Db 713 VESSTPQLPWRKAIWNKFKHNLVLEKRLLEKESTSTGSDSESTTTVVAENIDE 772
QY 770 -TSTTSDKPAPTTPKETAPTTPKPAPTTPKPAPTTPE---TPPTTSE-VSTPTTKEP 825
Db 773 VTTTEKRVQTPPTTBEKSTTQBEETTTTTTTEKTTSTKTTTEKPTTSESAITTEITSEP 832
QY 826 TTIHKSPESTPELSAETPKALENSPKPGVPTTKTPAATKPEMTTAKDKTTERDLRT 885
Db 833 ST-----TEST-----IVDTSSAITESSSTAETTTISAE--- 862
QY 886 TPETTTA-----APKMKETATTTEK 906
Db 863 TSETTSSAFTIGESPENTALOSSOKSEENESSAEKPGARDFVFKKHITTVKPAET 922
QY 907 TTESKITATTQ-VTST---TTQDTPPKITTLATTLAPKVT---TTKKTITTEIMNK 959
Db 923 TSAAVASTTTTTEPTTTEKSTTLETTPIEATTLNEVTPGAVTGPVDETTINTLELLSK 982
QY 960 PEETAKPDRAATNSKATTPKQO-----KPTAKPKPTSTKXKPTKTPRV 1002
Db 983 -----TNTQISQPKPTD:SKTDALSSLSLIGSFTKAPMAPTI----- 1022
QY 1003 RPKPTTTPKMTSTMPBLN-----PTSRIAEAMLOTTTRPNQTPNSKLVEN 1050
Db 1023 ---HTTTDAAFVTATEASLNDGSDKKIIDEAQPTDEIRA-----OPTN----- 1063
QY 1051 PKSEDAGGAEGETPHMLLRPHVFWPEVTPDMDYLPVRPNQGI 1092
Db 1064 -----EMDKEMEFEKRIREQRI 1080
```

```
RESULT 14
A48666
cell proliferation antigen Ki-67, long form - human
C:Species: Homo sapiens (man)
C:Date: 03-May-1994 #sequence_revision 03-May-1994 #text_change 09-Jul-2004
C:Accession: A48666
R:Schlueter, C.; Duchrow, M.; Wohlenberg, C.; Becker, M.H.G.; Key, G.; Flad, H.D.; Gerdes
J. Cell Biol. 123, 513-522, 1993
A:Title: The cell proliferation-associated antigen of antibody Ki-67: a very large, ubiqui-
ins.
A:Reference number: A48666; MUID:94043435; PMID:8227122
A:Accession: A48666
A>Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-3256 <SCH>
A:Cross-references: UNIPROT:P46013; EMBL:X65550; NID:G415818; PIDN:CAA46519.1; PID:G41581
C:Keywords: alternative splicing; cell cycle control; nucleus; tandem repeat
F:29-91/Domain: kinase interaction domain homology <KIH>

Query Match 14.8%; Score 863.5; DB 2; Length 3256;
Best Local Similarity 16.1%; Pred. No. 1.3e-09;
Matches 400; Conservative 137; Mismatches 447; Indels 1503; Gaps 93;

QY 22 VSGQELSCK-----GRCPESEFGRGECDCD 46
Db 741 ISNQKDFKEDLSGIAEMFKTPVKBQPLTSTCHIAISNENLLGKQFGTDSGEE--- 796
QY 47 AOCKKYDKCCPDYEGFCAEV---HNPTSPSSKKAPPPSGASQTIKSTTKRSPKPPNKK 102
Db 797 -----PLLPSTSEFGNVFSAQNAAKQPSDKCSASPLLRQCIRENGNVAKTFRNTY 849
QY 103 KTKKVIESIEEHSVSNQSSSSSSSSSSSSSSSIWIKIKSSKNSAANRELQK----- 155
Db 850 KMTSL-----ETKSDTETEPKTVSVNRSGRTEFRNTQKLPVESKS 893
QY 156 -----KLKVKDN-----KKKRT 167
Db 894 EBTNTEIVEICILKRGQKATLLQORREGEMKEIETPFYKNIENIELKENDERKMKAMKRSRT 953
QY 168 -KKKTPPKPVVDVDEAGSLDNGDFKVTTPDT-----STTOHNKVSTSPKITT 213
Db 954 WQKCAPMSDLTD-----LKSLPDTELMKDTARGNLLQTDQAKAPKSEKGI 1002
QY 214 AKPINRPSLPP---NSDTS-----KETSITVKNKETTVEKTTTNTKOTS 256
Db 1003 TK--NPKQSLQPEPINTPHTTKQLKASLGKVGKVEELLAVGKFRT-SGETTTHREPA 1059
QY 257 TTQKE-----KTTSAKETQSIEKTSADKLAPTSKVLAKPTP 292
Db 1060 GDGKSIRTPKSPKQILDPAARVTGMKWPRTPKPEAQSLE-----DLAGKELFQTPGP 1114
QY 293 KAE-----TTTK-----GPAL 303
Db 1115 SEESMTDEXTTKIACKSPPPSVDTPTSTKQWPKSLKADVEEBFLARKLTPSAGAM 1174
QY 304 TTPKEP-----TPT-----TPKEPAST----- 320
Db 1175 LTPKAGGDKDIKAFMGTPVQKLDLAGTLPGSKRQLQTPKEKAQALEDLAGKELFQTP 1234
QY 321 -----TPKEP-----TPTTIKAPT----- 335
Db 1235 GTEELVAAGKTKTIPCDSPQSDPVDVTSTKQRPKRSIRKADVEGELLACENLMPKAGK 1294
QY 336 ---TPKEPAP-----TTTKSAPTTTPKEPAPT----- 358
Db 1295 AMHTPK---PSVGEKDIIFVGTVPVQKLDLLENLTGSKRRPQTPKEKAQALEDLTGPK 1351
QY 359 -----TTKEPAPTTTPKEPAPTTPKEPAPTTPKSAPTTPKE----- 393
Db 1352 LFQTPGHTTEAFAAGKTTTKMCPSESPESADT-----PTSTRQPKTPLEKRDVQKELS 1405
QY 394 -----PAPTTPKPAPTTPKEPA--- 411
```


Db 1406 ALKKLTQTSGETHTDVKVGGEDKSIKAFRETAQKQKLPAAASVTGSKRHKPT-KEKAQPL 1464
QY 412 -----PTTKP-----EPTPT-TPKEPAPTTKE----- 432
Db 1465 EDLAGNKKLFQTPVCTDXTTETTKTKIACRSQPDVDTPTSSKQSKSLRKVDVEEF 1524
QY 433 -----PART-----TPKEPA-- 442
Db 1525 FALRKRTPSAGKAMHTPKPAVSGEKNIYAFMGTPVQKLDLTENLTGSKERLQTPKEQA 1584
QY 443 -----PTAPKPA----- 450
Db 1585 LEDLAGFKELFQTRGHTBESMINDKATAKACKSSQPDLDKXNFKSRRLKTSLGKGVKE 1644
QY 451 -----PTTPKEPAPT-----TPKEPA 466
Db 1645 ELLAVGKLTQTSGETHTTETPTGDKSMKAFMESPKQILDGAASLTGSKRQLRTPKGS 1704
QY 467 -----PTTTKEPSPTPK-----BP-----APTTPKSA- 490
Db 1705 EUPEDLAGFIELFQTPSHYKE-SMTNEKTKVSYRASQPDLDVDTPTSSKQPKRSLRAD 1763
QY 491 -----TTTK 494
Db 1764 TBEFLAFKQTPSAGKAMHTPKPAVGEKQDINTFLGTPVQKLDOPGNLPGSNRRLOTRK 1823
QY 495 EPA-----PT-----SPTTTPKEPSPTTKE----- 518
Db 1824 EKAQALELTGTFRELFQTPCTDNPTADEKTKKILCKSPQSDPAPTPTNKORPKRSLKK 1883
QY 519 -----PA-----PTT 523
Db 1884 ADVEEFLAFKLTSPSAGKAMHTPKAAVGEKQDINTFVGTPEVKDLDLGNLPGSKRRPQT 1943
QY 524 PKEPAPT-----TP-----KKPAPTTPKEPA----- 544
Db 1944 PREKAKALEDLAGFKELFQTPGHTBESMTDDKITEVSCSPQPDVKTPTSSKQBLKLSL 2003
QY 545 -----P 545
Db 2004 GKVGKVEVLVPGKLTQTSGETHTQHRETAGDGKSIKAFKESAKQWLPANVGTGWERWP 2063
QY 546 TTPKEPA-----PTTKKAPAPTAPKEPAP-----TTPKET--APT 579
Db 2064 RTPKEBAQSLDLAGFKELFQTPDHTTEESTTDDTKTKIACKSPPPESMDTPTSTRRPKT 2123
QY 580 P-----KKLTPTT-----PEKLAP-----TTPKEPAPT 603
Db 2124 PLGKRDIVEELGALKQLTQTTHTDKVGDDEKGINVRETAKQKLDPAASVTGSKRQRT 2183
QY 604 P-----PELA-----PTTPBEPPTTPBEPAPT----- 626
Db 2184 PKGKAQPLEDLAGFKELFQTPVCTDKPTTETTKIACRSQPDVGTPTIFKQSKRSL 2243
QY 627 -----TPKAAA-----P 633
Db 2244 RKADVEESLALRKRTPSVGKAMDTPKPAGGDEKQKAFMGTPVQKLDLPGNLPGSKRWP 2303
QY 634 NTPKEPA-----PTTPKEPAPTTPKEPAP-----TTPKET----- 663
Db 2304 QTFKEKAQALEDLAGFKELFQTPGTDKPTTDEKTKIACKSPQPDVDTPTASTKORPKRN 2363
QY 664 -----APTPK-----GTAPTTLKE 678
Db 2364 LRKADVEEFLALRKRTPSAGKAMDTPKPAVSDKRNINTFVETPVQKLDLGNLPGSKRQ 2423
QY 679 PAPTTPKPKAP-----KELAPTTTKEPTSTSDK-----PAP----- 710
Db 2424 ---POTPKKAEALEDLVGFKELFQTPGHTBESMTDDKITEVSCSPQESFKTSRSSQR 2481
QY 711 -----TTPKGTAPT-----TPK--EPAPT--- 727

Db 2482 LKIPLVKVDKBEPLAVSKLRTSGETTQTHTEPTGDSIKAFKESPKQILDPAASVTG 2541
QY 728 -----TPKEPA-----PTTPKGTAPTTL 745
Db 2542 SRRLRTRKEKARALEDLVDFKELFSAPGHTBESMTIDKNTKIPCKSPPPDELDTATSTK 2601
QY 746 KEPAPTTPKPAKELAP-----TTTKGPTSTSDKA----- 778
Db 2602 RCP-KTRPRKEVKEELSVAVERLTQTSQSTHTHKEPASGDEGKVLKQRAKKPNPVEEE 2660
QY 779 -----PTTPKE-----TAPTPKEPAPTTPKPAPTTPTPET 808
Db 2661 PSRRPRAPKEKAQPLEDLAGFTSELSETSGHTQESLTAGKATKIPCESPPLVVDITAST 2720
QY 809 PPPTTSVSPTTKETTKTHKSPDESTPELSAEP-TP-----KALENSPKPGVPTT- 860
Db 2721 KRHLRTRVQVQVKEEPSAV-KFTQTSGETTDAKGEKPAKESAKOTPPAPAS 2779
QY 861 -----KTPAA--TKPEMTTAKDKTTERDLRTPE---TTT 891
Db 2780 VTGSRPRAPRESAQAIEDLAGFKDPAAGHTEESMT---DDKTKIPCKSSPELEDAT 2836
QY 892 AAPKMTKETATTTBKTE---SKITATTTQVSTTTQ-----DTTPFKITILKTTTLA 941
Db 2837 SSKERPRTRAQKVEKSELLAVGKLTQTSGETTHTDKEPVGEGKGTAKFK----- 2886
QY 942 PKVTTTKTITTTIMNKPBEETAKPKDRA-----T 971
Db 2887 ---QPAKRVDAEDVIGSRQPRAPKEKAQPLEDLASQELSQTPGHTTELANGAADSFT 2943
QY 972 NSKATTPKPKPTK-----APK-KPT----- 991
Db 2944 SAPKQTPDSGKPLKISRRLRAPKVEPVGDVWVSTRDPVKSQSKSNTSLPLPFKRGCGKD 3003
QY 992 -----STKKPKTMRVRKEKTTPT---PKMTSTWPELNPTS 1025
Db 3004 GSVTGTGVRKLCMPAPESIVEELFASKQVAPRARGKSEPVVIMKSLTSKRIEPAE 3063
QY 1026 RIABAMLQTTTRPNQ-----TPNSKLVEVNPKESDAGAGETPHML------R 1069
Db 3064 ELNSDMKTNKEHKQLDSVPENKGISLRSRQDKTEAEQOITEVFLAEIRIEINRNEKK 3123
QY 1070 PHVEMPEVTDMMDYLPRVFNQGIINP 1096
Db 3124 P-----MKTSPEMD-----IQNP 3136

RESULT 15
Tl8535
high molecular mass nuclear antigen - chicken (fragment)
C:Species: Gallus gallus (Chicken)
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 09-Jul-2004
C:Accession: Tl8535
R:Shimada, K.; Hazata, M.; Mizuno, S.
J. Cell Sci. 110, 3031-3041, 1997
A:Title: A nuclear matrix-associated high molecular mass nuclear antigen, HMNA, of chick
A:Reference number: Z18955; MUID:9803440; PMID:9365273
A:Accession: Tl8535
A:Status: preliminary; translated from GE/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-1151 <SH1>
A:Cross-references: UNIPROT:O57580; EMBL:D88440; NID:d1177138; PID:d1025045; PIDN:BAA241

Query Match 14.8%; Score 860.7; DB 2; Length 1151;
Best Local Similarity 22.2%; Pred. No. 3.5e-10;
Matches 298; Conservative 89; Mismatches 360; Indels 593; Gaps 58;

QY 69 PTSPSSK-----KAPPPSG-----ASQTI 88
Db 27 PTAPRKPWPIAELHPAAQPPKWPVIGAPPPPTETPTPSKPTDGADAAPKASAE 86
QY 89 KSTTKRSPKPKTKTKVIESEITEHSVSENQESSSSSSSSSSSTTWIKSKNSA 148

```
Db 87 TSPPPASPPDPGPK-----APSGAGEAAGT----- 113
QY 149 ANRELQKLVKXDNKKRKKKPTK-----PPVDEAG-----SGLDNG-DF 190
Db 114 -----PPSQGAGTTPPSOGAAGPKGDGTAAQPSGKSGADG 151
QY 191 KVTTPDTSTTOHNVKSTGPKITTAKEPINSRPLPNSDTSKETSLTVNKETTVETKETTT 250
Db 152 KPAQDV-----PKAITA-ATEAR-----ASAAPTVPKATA----- 183
QY 251 TNKQSTSDGKXKTTSAKETQSTIEK-----TSAKDLAPTSKVLAKP----- 290
Db 184 -----EATAVTAASQSAKPAATDAATAASQSAKPAATVEVEKFAAAVAKEAKAV 233
QY 291 -----TPKAETTTKGPALTTKPEP-----TPTTKEPASTTTPKEPTTTIKSA 333
Db 234 TAAAAAPKATAEAKPAPVTSPTIPCSSAEAKPLTAASPTASKATAEAKPVPATASLAKTK 293
QY 334 PTTKPEPAPTTTTSKAPTTPKEPAPTTK-----EPAPTTTKE 370
Db 294 VTAEAKPAP-----SPSVPK-----ATTDKAVTATAPKAGPDVKPVAVCAEAKPAPPPPPQ 346
QY 371 PAPTTTKEPAPTTTTSKAPT-----PKEPA-----PTTPKKPAPTTTKEPAPTTTKE 417
Db 347 QLPKAAAAAAPTGTLEKPAATPHGSPRANGHTVTPPNVPRAAAATVP--TAGAVPKA 404
QY 418 PTTTTPKPEPAPTTKPEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTT 477
Db 405 STGTTT-----AAAPQOQV-----KAAVTPPSPOQAVPR-----AATAAAAPVT 445
QY 478 PKEPAPTTTTSKAPTTPKEPAP-----TTTKSAPTTPKEPSPT--ITKKEPAPTTTKEP- 527
Db 446 POOP---VTKAATTTNATPPPOPIPKAATTTTATVTPQOPIPKAGTDAAPPAVPKAPS 502
QY 528 -----APTTPKKPAPTTPKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKE 566
Db 503 DGRAATPGVNDATDPQKPPPTFQSVSAVTEPKPEQRAAPPSPNEATPAVPSPNLKS 562
QY 567 PAPTTPK-----ETAPTTPKLUTTTTPEKLAAPTTPKEPAPTTTPEELA 608
Db 563 PLFTIPKPVPELMALTFQVTAQMTQLAATKPSPIVPKASPK-ALMTBPPPPPLPRALA 621
QY 609 -----PTTP-----EPTPTTP-----BEPAPTTPKAA----- 631
Db 622 AAKLGLSPSPVASAMHAKVTRPLPASPVPMAASPSASLGPDAAARVALATNAASPGAKPE 681
QY 632 ----- 631
Db 682 AAGNGTLMAPMAAANTOMAPIGAAGAACTAPMGAAATHVSPMGAGGATOMSPGAAANTH 741
QY 632 -----APNTKPEP-----APTTKPEP-----APTTKPEP----- 655
Db 742 MSPIGAGGATOMSPMGAAANTOMSPMGATTONKSPMGAAANTQSPMGAAANTQVATAGN 801
QY 656 -----APTTPKETAPTTPKGT---APTTLKEP--APTTPKKPAKELAPTTTKE 699
Db 802 TMQVSPMGAAANTPQTPSPVGAATTPQSPMGAAATTLMSMGAAATTPQ---PSPMGAVTTQP 858
QY 700 P--TSTTSDKAP-----TTPKGTAPTTKPEP--APTTPKPEPAPTTTKEPAPTTTKEPAP 749
Db 859 PMAATNTQPPPMMAASTPQSTPMGAATTTQSPMGATTTQSPMGATTTQSP-----PMGA- 905
QY 750 PTTPKKPAKELAPTTTKEPAPTTTSDKAPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAP 805
Db 906 -STQAP-----PIVAGSPt-----PPPIPPSPPTACTSPQMSKSPPPDPKAPSA 953
QY 806 PETPPPTTSEVSTPTTKEPTTI---HKSPDESTPELSAETPKALENSKPEPGVPTTKT 862
Db 954 QTSAPAAHVANASPGVTVASPAPIGVTEASPSADGARLS--EGPTAATDGPK-----AS 1004
QY 863 PAATKPEMTTAKDKTTERDLRTTPTTTTAAKMTKETTATTTTETKTTTATTTQVTTST 922
Db 1005 PAAT-ADVTEAATDVTA---AATAVPAEAAPTAKKSSSSSSSSSSSSSSSSSSSSSSSSSS 1059
```

```
QY 923 TTODTTPFKITTLKTTTTLAPKVTTTTKKTITTTTINMKPBETAKPKDRATNSKATTPKPK 982
Db 1060 SDSA-----SSSSSSSSSNPAS 1075
QY 983 PTKAPKKPTSTKKPKTMTPRVRKPTTTPTRKMTSTMPELNPTSRIAEAMLOTTTTRPNQTP 1042
Db 1076 PA-----PAVGDQOQOQMTPGAAQSVPP-----VTEAAVQ----- 1104
QY 1043 NSKLVEVNPKSEDAAGAEGE 1062
Db 1105 -----EAAAAAAGAEERE 1119
```

Search completed: October 13, 2004, 11:56:32
Job time : 64.3855 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: October 13, 2004, 11:23:49 ; Search time 87.8789 Seconds
(without alignments)
4449.477 Million cell updates/sec

Title: SEQ1-B
Perfect score: 5826
Sequence: 1 MAWKTLPIYLLLSVFIQ.....DMDYLPVFNQGIINPMLS 1090

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.1

Searched: 2002273 seqs, 358729299 residues

Total number of hits satisfying chosen parameters: 2002273

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A_Geneseq_23Sep04:.*
1: Geneseq1980s:.*
2: Geneseq1990s:.*
3: Geneseq2000s:.*
4: Geneseq2001s:.*
5: Geneseq2002s:.*
6: Geneseq2003as:.*
7: Geneseq2003bs:.*
8: Geneseq2004s:.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Match | Length | DB ID | Description |
|------------|--------|-------|--------|-------|--------------------|
| 1 | 5811 | 99.7 | 1299 | 4 | AAM24322 Human EST |
| 2 | 5811 | 99.7 | 1404 | 2 | AAR26049 MSF precu |
| 3 | 5811 | 99.7 | 1404 | 4 | AAB29773 Human meg |
| 4 | 5811 | 99.7 | 1404 | 4 | AAB60568 Human meg |
| 5 | 5811 | 99.7 | 1404 | 8 | ADM98014 Human meg |
| 6 | 5811 | 99.7 | 1415 | 4 | AAB32262 Novel hum |
| 7 | 5788 | 99.3 | 1404 | 7 | ADK65839 Angiogene |
| 8 | 5608.7 | 96.3 | 1311 | 8 | ADK67912 Human ext |
| 9 | 5370.2 | 92.2 | 1320 | 7 | ADK65819 Angiogene |
| 10 | 5344.6 | 91.7 | 1270 | 8 | ADK67911 Human ext |
| 11 | 2920 | 50.1 | 546 | 4 | ABU53252 Human tes |
| 12 | 2888 | 49.6 | 551 | 4 | ABU53253 Human tes |
| 13 | 2880.1 | 49.4 | 902 | 4 | ABU53254 Human tes |
| 14 | 2757 | 47.3 | 513 | 4 | ABU53254 Human tes |
| 15 | 1994.2 | 34.2 | 452 | 2 | AAR80041 Human meg |
| 16 | 1409.9 | 24.2 | 538 | 5 | AAO18834 3' cartil |
| 17 | 1394.5 | 23.9 | 5179 | 6 | ABP55365 Human col |
| 18 | 1394.5 | 23.9 | 5179 | 6 | ABO07258 Human p53 |
| 19 | 1394.5 | 23.9 | 5179 | 7 | ADD48091 Human Pro |
| 20 | 1394.5 | 23.9 | 5179 | 7 | ADD44998 Human Pro |
| 21 | 1394.5 | 23.9 | 5179 | 7 | ADD48091 Human Pro |
| 22 | 1335.4 | 22.9 | 292 | 5 | AAU11261 Human HAP |
| 23 | 1127.2 | 19.3 | 1664 | 2 | AAW43106 C. thermo |
| 24 | 1103.7 | 18.9 | 1795 | 4 | ABE69806 Drosophil |
| 25 | 1099.6 | 18.9 | 717 | 4 | ABU53144 Human tes |

ALIGNMENTS

RESULT 1
AAM24322
ID AAM24322 standard; protein; 1299 AA.
XX AC AAM24322;
XX AC AAM24322;
DT 12-OCT-2001 (first entry)
XX DE Human EST encoded protein SEQ ID NO: 1847.
XX KW Human; sheep; pig; cow; fruit fly; Yeast; hamster; macaque; horse;
KW tomato; monkey; dog; sea urchin; expressed sequence tag; EST;
KW diagnostics; forensic test; gene mapping; genetic disorder; biodiversity;
KW gene therapy; nutrition.
XX OS Homo sapiens.
XX PN WO200154477-A2.
XX PD 02-AUG-2001.
XX PF 25-JAN-2001; 2001WO-US002687.
XX PR 25-JAN-2000; 2000US-00491404.
XX PR 17-JUL-2000; 2000US-00617746.
XX PR 03-AUG-2000; 2000US-00631451.
XX PR 15-SEP-2000; 2000US-00663870.
XX PA (HYSE-) HYSEQ INC.
XX PI Tang YT, Liu C, Zhou P, Qian XB, Wang Z, Chen R, Asundi V;
XX PI Cao Y, Drmanac RA, Zhang J, Werhman T;
XX DR WPI; 2001-476164/51.
XX DR N-PSDB; AAH98981.
XX PT Isolated polypeptide for treatment of diseases, diagnostics, raising
XX PT antibodies and research use.
XX PS Claim 20; Page 1198-1201; 1275pp; English.
XX CC The present invention provides the protein and coding sequences of novel
XX CC proteins from a variety of organisms, including human, dog, cat, horse,
XX CC cow, pig, hamster, monkey, macaque, yeast, bacteria, fruit fly, sea
XX CC urchin and tomato. These were derived from expressed sequence tags (ESTs)
XX CC from the organism of interest. They can be used in diagnostics,
XX CC forensics, gene mapping, identification of mutations, to assess
XX CC biodiversity and for nutritional purposes. The present sequence is a

26 1097 18.8 214 4 ABUS3255 Human tes
27 1091.9 18.7 8991 6 ABUS3255 Human tes
28 1053.7 18.1 763 3 ABUS3255 Human tes
29 1039.7 17.8 1049 4 ABUS3255 Human tes
30 1031 17.7 2284 4 ABUS3255 Human tes
31 1027 17.6 5703 8 ABUS3255 Human tes
32 1010.7 17.3 188 5 ABUS3255 Human tes
33 993.1 17.0 2112 4 ABUS3255 Human tes
34 987.9 17.0 770 4 ABUS3255 Human tes
35 987.1 16.9 4315 5 ABUS3255 Human tes
36 981 16.8 778 4 ABUS3255 Human tes
37 974.4 16.7 692 4 ABUS3255 Human tes
38 964.1 16.5 717 4 ABUS3255 Human tes
39 964.1 16.5 717 4 ABUS3255 Human tes
40 964.1 16.5 717 4 ABUS3255 Human tes
41 964.1 16.5 717 4 ABUS3255 Human tes
42 964.1 16.5 717 4 ABUS3255 Human tes
43 964.1 16.5 717 4 ABUS3255 Human tes
44 964.1 16.5 717 4 ABUS3255 Human tes
45 960.8 16.5 1460 7 ABUS3255 Human tes

XX PS Claim 1, 2 and 3; Fig 1; 87pp; English.

CC The sequence given is a full length translation from the megakaryocyte

CC stimulating factor (MSF) precursor. The sequence covered by exons II, III

CC and IV encodes megakaryocyte stimulating factor (MSF). This sequence is

CC modified by the addition of an N-terminal sequence encoding a secretory

CC leader, an initiating methionine preceding exon II and a terminating

CC codon following exon IV. The cDNA sequence given contains sequences

CC derived from human megakaryocyte colony stimulating factor (meg-CSF).

CC Exon I contains the initiating methionine, and encodes a classical

CC mammalian protein secretion signal sequence. The sequence encoding the

CC original meg-CSF includes exons II-IV and is thought to terminate in the

CC region between amino acid residues 134 - 147. The primary transcript of

CC this gene may be cleaved in different ways to yield a family of mRNA's

CC each encoding a different MSF protein. Exons V and VI are thought to be

CC related to the activity of the factor and are also implicated in the

CC stability, folding and processing of the molecule. These exons are also

CC thought to play a role in the observed synergy of MSF with other

CC cytokines. Exons V - XII are believed to be implicated in the processing

CC or folding of the appropriate structure of the resulting factor, i.e. one

CC or more of these exons may contain sequences which direct proteolytic

CC cleavage, adhesion, organisation of the cellular matrix or extracellular

CC matrix processing. Both naturally occurring and non-naturally occurring

CC MSF's may be characterised by various combinations of alternatively

CC spliced exons from this sequence, with the exons spliced together in

CC differing orders to form different members of the MSF family. (Updated on

CC 25-MAR-2003 to correct FN field.)

XX SQ Sequence 1404 AA;

Query Match 99.7%; Score 5811; DB 2; Length 1404;

Best Local Similarity 95.6%; Pred No. 6.8e-159;

Matches 1090; Conservative 0; Mismatches 0; Indels 50; Gaps 1;

QY 1 MAWKTLPIVYLLLSVFIQVSSQDLSSCAGRCGEGYSDATCNCYNQCQHYMECCPDF 60

DB 1 MAWKTLPIVYLLLSVFIQVSSQDLSSCAGRCGEGYSDATCNCYNQCQHYMECCPDF 60

QY 61 KVCVTAELSCKGRCFESFERGECDDAOCKYKDYCCPDYESFCA----- 105

DB 61 KVCVTAELSCKGRCFESFERGECDDAOCKYKDYCCPDYESFCAEVNHNPTSPPSKKAP 120

QY 106 -----EEHSVSNQBSSSSSSSSSSSSSSIW 130

DB 121 PPSGASQTKISITKRSPPKPKKTKKVIIEEHSVSNQBSSSSSSSSSSSSIW 180

QY 131 KIKSSNSAANRELQKLVKONKQRTKKTPPPVVDVDEAGSLDNGDFKVTPTDST 190

DB 181 KIKSSNSAANRELQKLVKONKQRTKKTPPPVVDVDEAGSLDNGDFKVTPTDST 240

QY 191 TOHNKYVSTGPKITTAKPINPRSLPNSDTSKETSLSLVNKETVETKETTINKGTSTDG 250

DB 241 TOHNKYVSTGPKITTAKPINPRSLPNSDTSKETSLSLVNKETVETKETTINKGTSTDG 300

QY 251 KEKTTSAKTSQTSIEKTSKADLAPTSKVLAKPTPKAETTTKGPAITTPKEPTTTPKEPAS 310

DB 301 KEKTTSAKTSQTSIEKTSKADLAPTSKVLAKPTPKAETTTKGPAITTPKEPTTTPKEPAS 360

QY 311 TTPKEPTPTTIKSAPTTTPKEPAPTTTKSAPTTTPKEPAPTTTPKEPAPTTTPKEP 370

DB 361 TTPKEPTPTTIKSAPTTTPKEPAPTTTKSAPTTTPKEPAPTTTPKEPAPTTTPKEP 420

QY 371 APTTTKSAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTT 430

DB 421 APTTTKSAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTT 480

QY 431 EPAPTAPKPAPTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTT 490

DB 481 EPAPTAPKPAPTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTT 540

QY 491 TTKSAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPK 550

DB 541 TTKSAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTKP 600

QY 551 APTAPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTT 610

DB 601 APTAPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTT 660

QY 611 PEPAPTTPKAAPNTPKPEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTT 670

DB 661 PEPAPTTPKAAPNTPKPEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTT 720

QY 671 APTTPKPAKELAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTT 730

DB 721 APTTPKPAKELAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTT 780

QY 731 TAPTTLKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPK 790

DB 781 TAPTTLKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPK 840

QY 791 KPAPTTTPETPPPTTSEVSTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPK 850

DB 841 KPAPTTTPETPPPTTSEVSTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPK 900

QY 851 TKTPAAATKPEMTTAKDKTTERDLRTPPETTTTAAAPKXMTKETTATTEKTESKITATTQV 910

DB 901 TKTPAAATKPEMTTAKDKTTERDLRTPPETTTTAAAPKXMTKETTATTEKTESKITATTQV 960

QY 911 TSSTTTQDTPPKITTLTKTTLAPKVTTKTITTTTEIMNKPEETAKPKDRATNSKATTPK 970

DB 961 TSSTTTQDTPPKITTLTKTTLAPKVTTKTITTTTEIMNKPEETAKPKDRATNSKATTPK 1020

QY 971 PQKETAPKXPTSTKPKTMRVRKPKTTPTPRMTSTMPELNPTSSIAEAMVLTQTTTRPN 1030

DB 1021 PQKETAPKXPTSTKPKTMRVRKPKTTPTPRMTSTMPELNPTSSIAEAMVLTQTTTRPN 1080

QY 1031 QTPNSKLVEVNPXSDEAGGAGETPHMLRPHVFMPEVTPDMOYLPRVNOGIIINPMLS 1090

DB 1081 QTPNSKLVEVNPXSDEAGGAGETPHMLRPHVFMPEVTPDMOYLPRVNOGIIINPMLS 1140

RESULT 3

AAB29773

ID AAB29773 standard; protein; 1404 AA.

XX AC AAB29773;

XX DT 28-FEB-2001 (first entry)

XX DE Human megakaryocyte stimulating factor (MSF), SEQ ID NO:1.

XX KW Human MSF; megakaryocyte stimulating factor; tribonectin;

XX KW alternative splicing; joint boundary lubricant; O-linked oligosaccharide;

XX KW osteoarthritis; tribosupplementation; tissue adhesion inhibition;

XX KW friction coefficient reduction; gene therapy; antiarthritic; osteopathic.

XX OS Homo sapiens.

XX PN WO200064930-A2.

XX PD 02-NOV-2000.

XX PF 24-APR-2000; 2000WO-US010953.

XX PR 23-APR-1999; 99US-00298970.

XX PA (RHOD-) RHODE ISLAND HOSPITAL LIFESPAN PARTNER.

XX PI Jay GD;

XX XX WPI; 2001-024673/03.

XX DR N-PSDB; AAC81498.

XX PT Novel tribonectin polypeptide useful as lubricant for treating

PT osteoarthritis, comprises O-linked lubricating moiety.

XX PS Claim 3; Page 7; 47pp; English.

CC The invention relates to a human tribonectin which is a product of

CC alternative splicing of the human MSF (megakaryocyte stimulating factor)

CC gene. The tribonectin has at least one O-linked oligosaccharide

CC lubricating moiety and has a polypeptide sequence comprising 1-76 repeats

CC of a motif having at least 50% identity to the sequence KEPAPTT

CC (AAB29774). The invention also relates to a nucleic acid encoding a human

CC MSF-derived tribonectin; a biocompatible composition comprising a human

CC tribonectin for inhibiting tissue adhesion formation; and a method of

CC diagnosing osteoarthritis or a predisposition to osteoarthritis by

CC measuring the amount of MSF or its fragment in a biological sample of a

CC mammal, wherein an increased amount of MSF compared to a control

CC indicates the presence of or predisposition to developing osteoarthritis.

CC The tribonectin and DNA encoding it are useful in the treatment of

CC osteoarthritis, where they may be used for lubricating mammalian joints,

CC such as articulating joints of humans, dogs or horses. The tribonectin,

CC when formulated as a membrane, foam, gel or fibre, is useful for

CC inhibiting adhesion between two surfaces such as the injured tissues of a

CC mammal, where the injury is caused by a surgical insertion or trauma, or

CC an artificial device e.g., an orthopaedic implant. In particular, one of

CC the surfaces is pericardial tissue. DNA encoding a tribonectin may be

CC used in gene therapy. The present sequence represents human MSF

XX XX

XX Sequence 1404 AA;

Query Match 99.7%; Score 5811; DB 4; Length 1404;

Best Local Similarity 95.6%; Pred. No. 6.8e-159;

Matches 1090; Conservative 0; Mismatches 0; Indels 50; Gaps 1;

QY 1 MAWKTLPIYLLLLSVFVIQVSSQDLSSCAGRCGEGYSRDATCNCYNCQHYNECCPDF 60

DB 1 MAWKTLPIYLLLLSVFVIQVSSQDLSSCAGRCGEGYSRDATCNCYNCQHYNECCPDF 60

QY 61 KRVCTAELSCKGRCFESFERGECDCDAQCKYDKCCPDYBSFCA----- 105

DB 61 KRVCTAELSCKGRCFESFERGECDCDAQCKYDKCCPDYBSFCAEVHNTSPSSKKAP 120

QY 106 -----EHSVSENQESSSSSSSSSSSSSTI 130

DB 121 PPSGASQTIKSTTKRSPKPPNKKTKKVIIESEITEHSVSENQESSSSSSSSSTI 180

QY 131 KTKSSKNAANRELQKLVKDNKNRTKKPTKPPVVDAGSLDNGDFKVTTPDTST 190

DB 181 KTKSSKNAANRELQKLVKDNKNRTKKPTKPPVVDAGSLDNGDFKVTTPDTST 240

QY 191 TOHNVKSTSPKITTAKPINRPSLPNNSDTSKETSLTVNKEITVETKTTTNKQSTGD 250

DB 241 TOHNVKSTSPKITTAKPINRPSLPNNSDTSKETSLTVNKEITVETKTTTNKQSTGD 300

QY 251 KEKTTSAKETOSIEKTSKADLAPTSKVLAKEPTKAEITTKGPAULTTKEPTTTPKEPAS 310

DB 301 KEKTTSAKETOSIEKTSKADLAPTSKVLAKEPTKAEITTKGPAULTTKEPTTTPKEPAS 360

QY 311 TTPKEPTTTIKSAPTTTKEPAPTTTKSAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEP 370

DB 361 TTPKEPTTTIKSAPTTTKEPAPTTTKSAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEP 420

QY 371 APPTTKSAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTK 430

DB 421 APPTTKSAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTK 480

QY 431 EPAPTAPKKPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPT 490

DB 481 EPAPTAPKKPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPT 540

QY 491 TTKSAPTTTKEPSPPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEP 550

DB 541 TTKSAPTTTKEPSPPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEP 600

QY 551 APTAPKEPAPTTTKEPAPTTPPKKLTTPTEKLAFTTPEKPAFTTPEELAPTTPEEPTPTT 610

DB 601 APTAPKEPAPTTTKEPAPTTPPKKLTTPTEKLAFTTPEKPAFTTPEELAPTTPEEPTPTT 660

QY 611 PEEPAFTTAKAANPTPKPAFTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKGTAFTTLKEP 670

DB 661 PEEPAFTTAKAANPTPKPAFTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKGTAFTTLKEP 720

QY 671 APTTPKKPAKELAPTTTKEPTSTTSDDKPAFTTPKGTAFTTPKEPAPTTTPKEPAPTTPKG 730

DB 721 APTTPKKPAKELAPTTTKEPTSTTSDDKPAFTTPKGTAFTTPKEPAPTTTPKEPAPTTPKG 780

QY 731 TAPTTTKEPAPTTPKPAKELAPTTTKEPTSTTSDDKPAFTTPKGTAFTTPKEPAPTTTPK 790

DB 781 TAPTTTKEPAPTTPKPAKELAPTTTKEPTSTTSDDKPAFTTPKGTAFTTPKEPAPTTTPK 840

QY 791 KPAPTTTPEPTTSEVSTPTTTPKEPTTIHKSPDSESTPELSAETPKALENSPKPGVPT 850

DB 841 KPAPTTTPEPTTSEVSTPTTTPKEPTTIHKSPDSESTPELSAETPKALENSPKPGVPT 900

QY 851 TKTPAATKPEMTTAKDKTTERDLRTTPETTTAAAPMKETAATTTKTESKITATTQV 910

DB 901 TKTPAATKPEMTTAKDKTTERDLRTTPETTTAAAPMKETAATTTKTESKITATTQV 960

QY 911 TSTTTQDTPPFKITTLKTTTLAPKVTITTKTITTTIMNKPEETAKPKDRATNSKATTPK 970

DB 961 TSTTTQDTPPFKITTLKTTTLAPKVTITTKTITTTIMNKPEETAKPKDRATNSKATTPK 1020

QY 971 PQKPTKAPKPTSTKKPKTMPRVKPTTPTPKMTSTMPELNPTSRIAEMLOTTTRPN 1030

DB 1021 PQKPTKAPKPTSTKKPKTMPRVKPTTPTPKMTSTMPELNPTSRIAEMLOTTTRPN 1080

QY 1031 QTPNSKLVEVNPKSDAGAGETPHMLLRPHVPEVTPDMDYLPVPNOGIINPMLS 1090

DB 1081 QTPNSKLVEVNPKSDAGAGETPHMLLRPHVPEVTPDMDYLPVPNOGIINPMLS 1140

RESULT 4

AAB60568

ID AAB60568 standard; protein; 1404 AA.

XX AAB60568;

AC

XX

DT 27-APR-2001 (first entry)

XX

DE Human megakaryocyte stimulating factor (MSF, CACP).

XX

XX Human; CACP protein; camptodactyly-arthropathy-coxa vara-pericarditis;

XX MSF; megakaryocyte stimulating factor; synovial lubricant;

XX chromosome 1q25-31; osteoarthritis; joint lubrication; osteopathic;

XX antiarthritic.

OS Homo sapiens.

XX

XX WO200107068-A1.

PN

XX

PD 01-FEB-2001.

XX

XX 21-JUL-2000; 2000WO-US0200002.

XX

XX 23-JUL-1999; 99US-0145328P.

XX 19-JUL-2000; 2000US-00145328.

XX

XX (UYCA-) UNIV CASE WESTERN RESERVE.

XX

XX Warman ML;

PI

XX

XX WPI; 2001-182721/18.

XX

XX New composition comprising the camptodactyly-arthropathy-coxa vara-

XX pericarditis protein in combination with an anesthetic, useful for

XX treating osteoarthritis, or as lubricants of tissue and joints.

XX

XX Example 1; Page; 34pp; English.

XX

CC The invention relates to a method of treating osteoarthritis via the
CC administration of a composition comprising the camptodactyl-arthropathy-
CC coxa vara-pericarditis (CACP) protein, or portions of the CACP protein.
CC The composition may further comprise a local anaesthetic. The composition
CC of the invention may be administered via intra-articular or intravenous
CC injection. The human CACP protein is identified in the invention as being
CC megakaryocyte stimulating factor (MSF). The gene encoding CACP protein
CC (MSF) is located on chromosome 1q25-31, and mutations in this gene are
CC responsible for the heritable disorder camptodactyl-arthropathy-coxa
CC vara-pericarditis, in which patients have synovial hyperplasia without
CC evidence of inflammation. CACP protein (MSF) acts as a synovium
CC lubricant, and can be used to lubricate tissue and joints in the
CC treatment of osteoarthritis. The composition may be applied to reduce the
CC symptoms of osteoarthritis (e.g., joint pain, loss of range of movement
CC or joint damage). The present sequence represents human megakaryocyte
CC stimulating factor (MSF, CACP protein). Note: this sequence is not given
CC in its entirety in figure 4 of the specification, although a GenBank
CC accession number was given. This sequence was therefore obtained from
CC GenBank (U70316)
XX
SQ Sequence 1404 AA;

Query Match 99.7%; Score 5811; DB 4; Length 1404;
Best Local Similarity 95.6%; Pred. No. 6.8e-159;
Matches 1090; Conservative 0; Mismatches 0; Indels 50; Gaps 1;

QY 1 MAWKLPYLLLLSVFIQVSSODLSACGCGEGYSRATCNCYNCOHYMECCDPF 60
DB 1 MAWKLPYLLLLSVFIQVSSODLSACGCGEGYSRATCNCYNCOHYMECCDPF 60
QY 61 KRVCFAELSCGRCPESFERGECDCQAQCKYKCCPDYSEFCA----- 105
DB 61 KRVCFAELSCGRCPESFERGECDCQAQCKYKCCPDYSEFCAEVHNPTSPRKKAP 120
QY 106 -----EHSVGENOESSSSSSSSSSSSSSSIW 130
DB 121 PPSGASQTIKSTKSPKPPNKKVKVISEITEHSVENQESSSSSSSSSSSIW 180
QY 131 KIKSKNSAANRELQKLVKDNKNKTKKTPPPVVDKAGSLDNGDFKVTPTDST 190
DB 181 KIKSKNSAANRELQKLVKDNKNKTKKTPPPVVDKAGSLDNGDFKVTPTDST 240
QY 191 TQHKVSTSPKITTAKPINRPSLPPNSDTSKETSITVKNKETTVEKTTTNKQTSIDG 250
DB 241 TQHKVSTSPKITTAKPINRPSLPPNSDTSKETSITVKNKETTVEKTTTNKQTSIDG 300
QY 251 KEKTSKETSQIEKTSADLAPTSKVLAKPTPKAETTTKGPAITPKPTTKEPAS 310
DB 301 KEKTSKETSQIEKTSADLAPTSKVLAKPTPKAETTTKGPAITPKPTTKEPAS 360
QY 311 TTPKEPTPTTIKSAPTTKPEAPTTTKSAPTTKPEAPTTTKPEAPTTKPEAPTTKEP 370
DB 361 TTPKEPTPTTIKSAPTTKPEAPTTTKSAPTTKPEAPTTTKPEAPTTKPEAPTTKEP 420
QY 371 APTTKSAPTTKPEAPTTKPEAPTTKPEAPTTKPEAPTTKPEAPTTKPEAPTTK 430
DB 421 APTTKSAPTTKPEAPTTKPEAPTTKPEAPTTKPEAPTTKPEAPTTKPEAPTTK 480
QY 431 EPAPTAPKKPAPTTKPEAPTTKPEAPTTKPEAPTTKPEAPTTTKSAPTTTKPEAPT 490
DB 481 EPAPTAPKKPAPTTKPEAPTTKPEAPTTTKESPPTTKPEAPTTTKSAPTTTKPEAPT 540
QY 491 TTKSAPTTKPEAPTTKPEAPTTKPEAPTTKPEAPTTKPEAPTTKPEAPTTTKKP 550
DB 541 TTKSAPTTKPEAPTTKPEAPTTKPEAPTTKPEAPTTKPEAPTTKPEAPTTTKKP 600
QY 551 APTAPKPEAPTTKPEAPTTKPEAPTTKPEAPTTKPEAPTTKPEAPTTPEEAPTTPEPTPTT 610
DB 601 APTAPKPEAPTTKPEAPTTKPEAPTTKPEAPTTKPEAPTTPEEAPTTPEPTPTT 660
QY 611 PEEAPTTKAAAPNTKPEAPTTKPEAPTTKPEAPTTKPEAPTTKPEAPTTKPEAPTTLKEP 670
DB 661 PEEAPTTKAAAPNTKPEAPTTKPEAPTTKPEAPTTKPEAPTTKPEAPTTKPEAPTTLKEP 720

QY 671 APTTKPKKPAKELAPTTTKPTSTSDKAPATPKCTAPTTPKEPAPTTPKEPAPTTPKG 730
DB 721 APTTKPKKPAKELAPTTTKPTSTSDKAPATPKCTAPTTPKEPAPTTPKEPAPTTPKG 780
QY 731 TAPTTLKEPAPTTPKPAKELAPTTTKGPTSTSDKAPATPKCTAPTTPKEPAPTTPK 790
DB 781 TAPTTLKEPAPTTPKPAKELAPTTTKGPTSTSDKAPATPKCTAPTTPKEPAPTTPK 840
QY 791 KPAPTTPPTPTTSEVSTPTTKPTTHKSPDESTPDSABPTPKALENSPKERGVP 850
DB 841 KPAPTTPPTPTTSEVSTPTTKPTTHKSPDESTPDSABPTPKALENSPKERGVP 900
QY 851 TKTPAAKTPMTTAKDKTTERDLRTPETTTAAAPKTKETATTTTKTTESKITTATTOV 910
DB 901 TKTPAAKTPMTTAKDKTTERDLRTPETTTAAAPKTKETATTTTKTTESKITTATTOV 960
QY 911 TSTTTQDTPFKITTLKTTTLAPKVTTKKTIITTTIMNKPEETAKPKORATNSKATTPK 970
DB 961 TSTTTQDTPFKITTLKTTTLAPKVTTKKTIITTTIMNKPEETAKPKORATNSKATTPK 1020
QY 971 POKPTKAPKPTSTKPKTMPRVKPKTTPTPEKMTSTMPELNPTSRIAEAMLOTTTRPN 1030
DB 1021 POKPTKAPKPTSTKPKTMPRVKPKTTPTPRMTSTMPELNPTSRIAEAMLOTTTRPN 1080
QY 1031 QTENSKLVEVNPKSEDAGGAEGETPHMLLRPHVMEPEVTPDMDYLRVFNQGIINPMLS 1090
DB 1081 QTENSKLVEVNPKSEDAGGAEGETPHMLLRPHVMEPEVTPDMDYLRVFNQGIINPMLS 1140
RESULT 5
ADM98014
ID ADM98014 standard; protein; 1404 AA.
XX
AC ADM98014;
XX
DT 01-JUL-2004 (first entry)
XX
DE Human megakaryocyte stimulating factor (MSF).
XX
KW lubricating polypeptide; O-linked oligosaccharide; joint lubrication;
KW CAP; camptodactyl-arthropathy pericarditis; osteoarthritis; human;
KW megakaryocyte stimulating factor; MSF.
XX
OS Homo sapiens.
XX
FN US2004072741-A1.
XX
ED 15-APR-2004.
XX
PF 02-JUL-2001; 2001US-00897188.
XX
PR 23-APR-1999; 99US-00298970.
PR 24-APR-2000; 2000US-00556246.
XX
PA (JAYG/) JAY G D.
XX
PI Jay GD;
XX
DR WPI; 2004-373948/35.
DR N-PSDB; ADM98015.
XX
PT New tribonectin polypeptides and polynucleotides for lubricating joints
PT or other tissues to prevent or treat Camptodactyl-arthropathy-
PT pericarditis syndrome or osteoarthritis.
XX
PS Claim 1; SEQ ID NO 1; 34pp; English.
XX
CC The invention relates to a lubricating polypeptide and at least one O-
CC linked oligosaccharide. The composition and methods are useful for
CC lubricating joints or other tissues to prevent or treat camptodactyl-
CC arthropathy pericarditis (CAP) or osteoarthritis in mammals. The present
CC sequence represents the amino acid sequence of the human megakaryocyte

| | | | |
|----|------|--|------|
| Db | 12 | MAWKTLPIYLLLLSVFVIQVSSQDLSSCAGRCGEGYSDATCNCYNOHYMECCPDF | 71 |
| Qy | 61 | KRVCTAELSCKRCFESFERGECDCDAQCKYDKCCPDYESFCA----- | 105 |
| Db | 72 | KRVCTAELSCKRCFESFERGECDCDAQCKYDKCCPDYESFCAEVNPTSPSSKKAP | 131 |
| Qy | 106 | -----EHSVSENQSSSSSSSSSSSTIW | 130 |
| Db | 132 | PPSGASQTIKSTTKRSPKPNKKTKVIESEITEEHSVSENQSSSSSSSSSTIW | 191 |
| Qy | 131 | KIKSSKNSAANELOKKLVKDKNKNRTKKKTPKPPVVDAGSLDNGDFKVTTPDST | 190 |
| Db | 192 | KIKSSKNSAANELOKKLVKDKNKNRTKKKTPKPPVVDAGSLDNGDFKVTTPDST | 251 |
| Qy | 191 | TOHNKVSTSPKITTAKPINRPSLPNSDTSKETSILTVNKETITVKETITTKNKTSDG | 250 |
| Db | 252 | TOHNKVSTSPKITTAKPINRPSLPNSDTSKETSILTVNKETITVKETITTKNKTSDG | 311 |
| Qy | 251 | KEKTTSAKTSQSIKTSKADLAPTSKVLAKPTPKAETTTKGPAITTPKEPTTPKEPAS | 310 |
| Db | 312 | KEKTTSAKTSQSIKTSKADLAPTSKVLAKPTPKAETTTKGPAITTPKEPTTPKEPAS | 371 |
| Qy | 311 | TTTKEPTPTTIKSAPTTKEPAPTTTKSAPTTKEPAPTTTKPAPTTKEPAPTTTKKEP | 370 |
| Db | 372 | TTTKEPTPTTIKSAPTTKEPAPTTTKSAPTTKEPAPTTTKPAPTTKEPAPTTTKKEP | 431 |
| Qy | 371 | APTTKSAPTTKEPAPTTKPKAPTTKPEPAPTTKPEPAPTTKPEPAPTTKPEAPTTPK | 430 |
| Db | 432 | APTTKSAPTTKEPAPTTKPKAPTTKPEPAPTTKPEPAPTTKPEPAPTTKPEAPTTPK | 491 |
| Qy | 431 | EPAPTAPKAPAPTTKEPAPTTKPEPAPTTKPEPAPTTKPEPAPTTTKSAPTTKEPAPT | 490 |
| Db | 492 | EPAPTAPKAPAPTTKEPAPTTKPEPAPTTKPEPAPTTKPEPAPTTTKSAPTTKEPAPT | 551 |
| Qy | 491 | TTKSAPTTKPEPAPTTTKPEPAPTTKPKAPTTKPEPAPTTKPEPAPTTTKKAP | 550 |
| Db | 552 | TTKSAPTTKPEPAPTTTKPEPAPTTKPKAPTTKPEPAPTTKPEPAPTTTKKAP | 611 |
| Qy | 551 | APTAPKEPAPTTKPEAPTTPKKLPTTPEKAPITPKAPAPTTPEELAPTTPEEPTTT | 610 |
| Db | 612 | APTAPKEPAPTTKPEAPTTPKKLPTTPEKAPITPKAPAPTTPEELAPTTPEEPTTT | 671 |
| Qy | 611 | PREPAPTTKAAAPNTPKEPAPTTKPEPAPTTKPEPAPTTKPEAPTTKGTAPTTLKEP | 670 |
| Db | 672 | PREPAPTTKAAAPNTPKEPAPTTKPEPAPTTKPEPAPTTKPEAPTTKGTAPTTLKEP | 731 |
| Qy | 671 | APTTPKAPKELAPTTTKEPTSTSDKAPAPTTKGTAPTTKPEPAPTTKPEAPTTKPKG | 730 |
| Db | 732 | APTTPKAPKELAPTTTKEPTSTSDKAPAPTTKGTAPTTKPEPAPTTKPEAPTTKPKG | 791 |
| Qy | 731 | TAPTTLKEPAPTTKPKAPKELAPTTTKEPTSTSDKAPAPTTKGTAPTTKPEPAPTTPK | 790 |
| Db | 792 | TAPTTLKEPAPTTKPKAPKELAPTTTKEPTSTSDKAPAPTTKGTAPTTKPEPAPTTPK | 851 |
| Qy | 791 | KPAPTTTPEPTTSEVSTPTTKEPTTIHKSPDESTPELSAETPKALENSPKFPGVPT | 850 |
| Db | 852 | KPAPTTTPEPTTSEVSTPTTKEPTTIHKSPDESTPELSAETPKALENSPKFPGVPT | 911 |
| Qy | 851 | TKTPAATKEMITTAQDKTERDLRTTPTTAAAPKMTKETAATTEKTESKITATTQV | 910 |
| Db | 912 | TKTPAATKEMITTAQDKTERDLRTTPTTAAAPKMTKETAATTEKTESKITATTQV | 971 |
| Qy | 911 | TSTTTQDTPPFKITTLLKTTTLAPKVTTTKITITTEIMNKPBEETAKPKDRATNSKATPK | 970 |
| Db | 972 | TSTTTQDTPPFKITTLLKTTTLAPKVTTTKITITTEIMNKPBEETAKPKDRATNSKATPK | 1031 |
| Qy | 971 | PQKPTAPKAPKPTSTKPKTMPVRKPKTPTPRKMTSTMPBLNPTSRIAEAMLOTTTRN | 1030 |
| Db | 1032 | PQKPTAPKAPKPTSTKPKTMPVRKPKTPTPRKMTSTMPBLNPTSRIAEAMLOTTTRN | 1091 |
| Qy | 1031 | QTPNSKLVEVNPXKSDAGGAGETPHMLLRPHVFMPEVTPDMDYLPVNPQGLIINPMLS | 1090 |

| | | | |
|----------|---|--|------|
| Db | 1092 | QTPNSKLVEVNPXKSDAGGAGETPHMLLRPHVFMPEVTPDMDYLPVNPQGLIINPMLS | 1151 |
| RESULT 7 | | | |
| ID | ADK65839 | standard; protein; 1404 AA. | |
| XX | ADK65839; | | |
| DT | 06-MAY-2004 | (first entry) | |
| XX | Angiogenesis-differentially expressed protein #53. | | |
| DE | cytostatic; cardiant; vasotropic; antiarteriosclerotic; | | |
| XX | angiogenesis inhibitor; angiogenesis stimulator; angiogenic index; | | |
| KW | gene expression; cancer; coronary artery disease; myocardial ischemia; | | |
| KW | coronary arteriosclerosis; forensic medicine. | | |
| OS | Homo sapiens. | | |
| XX | WO200306831-A2. | | |
| FN | 14-AUG-2003. | | |
| PD | 07-FEB-2003; 2003WO-US003848. | | |
| XX | 07-FEB-2002; 2002US-00067482. | | |
| PR | 10-JUN-2002; 2002US-00164595. | | |
| PR | 16-AUG-2002; 2002US-0403649P. | | |
| PR | 03-JAN-2003; 2003US-0437746P. | | |
| XX | (ORIG-) ORIGENE TECHNOLOGIES INC. | | |
| PA | Sun Z, Li X, Kovacs KF, Fan W, Jay G; | | |
| PI | WPI; 2003-731502/69. | | |
| XX | Determining the angiogenic index of a tissue or cell sample using | | |
| XX | expression levels of differentially expressed genes, useful for | | |
| PT | diagnosing or treating cancer, coronary artery disease, myocardial | | |
| PT | ischemia and/or arteriosclerosis. | | |
| XX | Disclosure; SEQ ID NO 78; 296pp; English. | | |
| PS | The invention relates to a method of determining the angiogenic index of | | |
| XX | a tissue or cell sample comprising assessing, in a sample, the expression | | |
| CC | levels of one or more differentially-expressed gene from any of 34 DNA | | |
| CC | sequences, given in the specification, where the levels are indicative of | | |
| CC | the angiogenic index. The methods and compositions of the present | | |
| CC | invention are useful for diagnosing, preventing and/or treating cancer, | | |
| CC | coronary artery disease, myocardial ischemia or coronary | | |
| CC | arteriosclerosis. They can also be used in research, drug discovery and | | |
| CC | forensic medicine involving angiogenesis. This sequence corresponds to | | |
| CC | one of the differentially expressed proteins of the invention. | | |
| XX | Sequence 1404 AA; | | |
| SQ | Query Match 99.3%; Score 5788; DB 7; Length 1404; | | |
| | Best Local Similarity 95.4%; Pred. No. 3.1e-158; | | |
| | Matches 1087; Conservative 0; Mismatches 3; Indels 50; Gaps 1; | | |
| Qy | 1 | MAWKTLPIYLLLLSVFVIQVSSQDLSSCAGRCGEGYSDATCNCYNOHYMECCPDF | 60 |
| Db | 1 | MAWKTLPIYLLLLSVFVIQVSSQDLSSCAGRCGEGYSDATCNCYNOHYMECCPDF | 60 |
| Qy | 61 | KRVCTAELSCKRCFESFERGECDCDAQCKYDKCCPDYESFCA----- | 105 |
| Db | 61 | KRVCTAELSCKRCFESFERGECDCDAQCKYDKCCPDYESFCAEVNPTSPSSKKAP | 120 |
| Qy | 106 | -----EHSVSENQSSSSSSSSSSSTIW | 130 |
| Db | 121 | PPSGASQTIKSTTKRSPKPNKKTKVIESEITEEHSVSENQSSSSSSSSSTIR | 180 |

QY 131 KIKSSKNSAANRELOKLVKDKKQNRKTKKPTKPPVVDKAGSLDNGDFKVTTPDST 190
 DB 181 KIKSSKNSAANRELOKLVKDKKQNRKTKKPTKPPVVDKAGSLDNGDFKVTTPDST 240
 QY 191 TQHNKYSTSPKLTAKPINRPSLSPNSDTSKETSITVKNKETTIVETKETTITNKQTSDDG 250
 DB 241 TQHNKYSTSPKLTAKPINRPSLSPNSDTSKETSITVKNKETTIVETKETTITNKQTSDDG 300
 QY 251 KEKTTISAKETQSIKTSKADLAPTSKVLAKPTPKAETTTKGPALTTPKEPTTPPKEPAS 310
 DB 301 KEKTTISAKETQSIKTSKADLAPTSKVLAKPTPKAETTTKGPALTTPKEPTTPPKEPAS 360
 QY 311 TPKKPTPTTIKSAPTTKPEAPPTTKSAPTTKBPAPTTKBPAPTTKBPAPTTKBP 370
 DB 361 TPKKPTPTTIKSAPTTKPEAPPTTKSAPTTKBPAPTTKBPAPTTKBPAPTTKBP 420
 QY 371 APTTTKSAPTTKBPAPTTKBPAPTTKBPAPTTKBPAPTTKBPAPTTKBPAPTTKBP 430
 DB 421 APTTTKSAPTTKBPAPTTKBPAPTTKBPAPTTKBPAPTTKBPAPTTKBPAPTTKBP 480
 QY 431 EPAPTAPKAPPTTKBPAPTTKBPAPTTKBPAPTTKBPAPTTKBPAPTTKBPAPTTKBP 490
 DB 481 EPAPTAPKAPPTTKBPAPTTKBPAPTTKBPAPTTKBPAPTTKBPAPTTKBPAPTTKBP 540
 QY 491 TTKSAPTTKBPAPTTKBPAPTTKBPAPTTKBPAPTTKBPAPTTKBPAPTTKBPAPTTKBP 550
 DB 541 TTKSAPTTKBPAPTTKBPAPTTKBPAPTTKBPAPTTKBPAPTTKBPAPTTKBPAPTTKBP 600
 QY 551 APTAPKBPAPTTKBPAPTTKBPAPTTKBPAPTTKBPAPTTKBPAPTTKBPAPTTKBPAPTTKBP 610
 DB 601 APTAPKBPAPTTKBPAPTTKBPAPTTKBPAPTTKBPAPTTKBPAPTTKBPAPTTKBPAPTTKBP 660
 QY 611 PEPAPPTPKAAPNTKBPAPTTKBPAPTTKBPAPTTKBPAPTTKBPAPTTKBPAPTTKBPAPTTKBP 670
 DB 661 PEPAPPTPKAAPNTKBPAPTTKBPAPTTKBPAPTTKBPAPTTKBPAPTTKBPAPTTKBPAPTTKBP 720
 QY 671 APTTPKBPAPKELAPTTKBPAPTTKBPAPTTKBPAPTTKBPAPTTKBPAPTTKBPAPTTKBPAPTTKBP 730
 DB 721 APTTPKBPAPKELAPTTKBPAPTTKBPAPTTKBPAPTTKBPAPTTKBPAPTTKBPAPTTKBPAPTTKBP 780
 QY 731 TAPTTLKBPAPTTKBPAPKELAPTTKBPAPTTKBPAPTTKBPAPTTKBPAPTTKBPAPTTKBPAPTTKBP 790
 DB 781 TAPTTLKBPAPTTKBPAPKELAPTTKBPAPTTKBPAPTTKBPAPTTKBPAPTTKBPAPTTKBPAPTTKBP 840
 QY 791 KPAPTTPTPTTSEVSTPTTKETPTTIHKSPDSTPELSAETPKALNSPKPGVPT 850
 DB 841 KPAPTTPTPTTSEVSTPTTKETPTTIHKSPDSTPELSAETPKALNSPKPGVPT 900
 QY 851 TKTPAATKPEMTTAKDKTTERDLRTTPTTAAAPKMTKETATTEKTESKITATTQV 910
 DB 901 TKTPAATKPEMTTAKDKTTERDLRTTPTTAAAPKMTKETATTEKTESKITATTQV 960
 QY 911 TSTTTQDTPPKITTLKTTTLAPKVTTKKTIITTEIMNKEETAKPKDRATNSKATTPK 970
 DB 961 TSTTTQDTPPKITTLKTTTLAPKVTTKKTIITTEIMNKEETAKPKDRATNSKATTPK 1020
 QY 971 FQKPTKAPKPTSTKPKTPMVRKPKTPTTPRKMSTMPNLNPTSRIAEAMLTQTRPN 1030
 DB 1021 FQKPTKAPKPTSTKPKTPMVRKPKTPTTPRKMSTMPNLNPTSRIAEAMLTQTRPN 1080
 QY 1031 QTPNSKLVNPKSEDAGAGETPHMLLRPHVMPVTPDMVLPVPCQGIINPMLS 1090
 DB 1081 QTPNSKLVNPKSEDAGAGETPHMLLRPHVMPVTPDMVLPVPCQGIINPMLS 1140

RESULT 8

ID ADK67912

XX ADK67912 standard; protein; 1311 AA.

AC ADK67912;

XX 06-MAY-2004 (first entry)

DT 06-MAY-2004 (first entry)

XX 06-MAY-2004 (first entry)

DE Human extracellular messenger (EXMES) polypeptide.
 KW Human, extracellular messenger; EXMES; respiratory-gen.; anti-allergic;
 KW antiasthmatic; anti-inflammatory; antidiabetic; neuroprotective;
 KW muscular-gen.; anarthritic; osteopathic; hepatotropic; antipsoriatic;
 KW virucide; fungicide; antiparasitic; protozoacide; antihelminthic;
 KW cytostatic; gene therapy.
 XX Homo sapiens.
 OS
 XX
 XX
 FH Key Location/Qualifiers
 FT Peptide 1..30
 FT /label= Signal_peptide
 FT /note= "Spans residues 1 to 18, 20, 21, 24, 29 or 30
 according to identification method"
 XX WO2004013292-A2.
 XX 12-FEB-2004.
 XX 30-JUL-2003; 2003WO-US024084.
 XX 02-AUG-2002; 2002US-0400810P.
 PR 19-SEP-2002; 2002US-0412197P.
 PR 04-OCT-2002; 2002US-0416004P.
 PR 08-NOV-2002; 2002US-0424862P.
 XX (INCY-) INCYTE CORP.
 PA Elliott VS, Khare R, Tran UK, Swarnakar A, Marquis JP,
 PI Richardson TW, Emerling BM, Lindquist EA, Chawla NK, Ramkumar J,
 PI Lee SY;
 XX WPI; 2004-157116/15.
 DR N-PSDB; ADK67917.
 XX New extracellular messengers and nucleic acids, useful for diagnosing,
 PT treating or preventing e.g. allergies, anemia, Crohn's disease, diabetes
 PT mellitus, myasthenia gravis, osteoporosis, psoriasis, cancers or
 PT autoimmune thyroiditis.
 XX Claim 60; SEQ ID NO 5; 165pp; English.
 XX The present sequence is that of novel human extracellular messenger
 CC (EXMES) Incyte ID NO: 7513018CD1 polypeptide. The protein shows homology
 CC to human megakaryocyte stimulating factor. The invention provides EXMES
 CC polynucleotides and polypeptides, as well as expression vectors, host
 CC cells, antibodies, agonists and antagonists, and methods for diagnosing,
 CC treating or preventing disorders associated with aberrant expression of
 CC EXMES, especially autoimmune and inflammatory disorders, cell
 CC proliferative disorders and endocrine disorders, e.g. adult respiratory
 CC distress syndrome, allergy, anaemia, erythroblastosis foetalis, Crohn's
 CC disease, diabetes mellitus, myasthenia gravis, osteoarthritis,
 CC osteoporosis, pancreatitis, psoriasis, viral, bacterial, fungal,
 CC parasitic, protozoal or helminthic infections, cancers, autoimmune
 CC thyroiditis, cretinism, Plummer's disease or thyroid carcinoma.
 CC Embodiments also provide methods for using the purified EXMES and/or
 CC their encoding polynucleotides for facilitating the drug discovery
 CC process, including determining of efficacy, dosage, toxicity and
 CC pharmacology, and for investigating the pathogenesis of diseases and
 CC medical conditions.
 XX Sequence 1311 AA;
 SQ
 Query Match 96.3%; Score 5608.7; DB 8; Length 1311;
 Best Local Similarity 96.1%; Pred. No. 4, 1e-153;
 Matches 1047; Conservative 0; Mismatches 0; Indels 43; Gaps 1;
 QY 1 MAWKTLPIYLLLLLSVFVIQQVSSQDLSSCAGRCGEGYSRDATNCNDYNCQHYMECCPDF 60
 DB 1 MAWKTLPIYLLLLLSVFVIQQVSSQDLSSCAGRCGEGYSRDATNCNDYNCQHYMECCPDF 60
 QY 61 KRVTCTAELSCKRCFESFERGECDCDAQCKYDKCCPDFYSCFAEEHVSSENQSSSS 120

ADK65819
ID ADK65819 standard; protein; 1320 AA.
XX
AC ADK65819;
XX
DT 06-MAY-2004 (first entry)
XX
DE Angiogenesis-differentially expressed protein ANH0316.
XX
KW cytosolic; cardiant; vasotropic; antiarteriosclerotic;
KW angiogenesis inhibitor; angiogenesis stimulator; angiogenic index;
KW gene expression; cancer; coronary artery disease; myocardial ischemia;
KW coronary arteriosclerosis; forensic medicine.
XX
OS Homo sapiens.
XX
FN WO2003066831-A2.
XX
PD 14-AUG-2003.
XX
XX 07-FEB-2003; 2003WO-US003848.
XX
XX 07-FEB-2002; 2002US-00067482.
XX 10-JUN-2002; 2002US-00164595.
XX 16-AUG-2002; 2002US-0403649P.
XX 03-JAN-2003; 2003US-0437746P.
XX
(ORIG-) ORIGENE TECHNOLOGIES INC.
PA Sun Z, Li X, Kovacs KP, Fan W, Jay G;
XX
XX WP; 2003-731502/69.
XX N-PSDB; ADK65818.
XX
PT Determining the angiogenic index of a tissue or cell sample using
PT expression levels of differentially expressed genes, useful for
PT diagnosing or treating cancer, coronary artery disease, myocardial
PT ischemia and/or arteriosclerosis.
XX
PS Claim 23; SEQ ID NO 58; 236pp: English.
XX
XX The invention relates to a method of determining the angiogenic index of
XX a tissue or cell sample comprising assessing, in a sample, the expression
XX levels of one or more differentially-expressed genes from any of 34 DNA
XX sequences, given in the specification, where the levels are indicative of
XX the angiogenic index. The methods and compositions of the present
XX invention are useful for diagnosing, preventing and/or treating cancer,
XX coronary artery disease, myocardial ischemia or coronary
XX arteriosclerosis. They can also be used in research, drug discovery and
XX forensic medicine involving angiogenesis. This sequence corresponds to
XX one of the differentially expressed proteins of the invention.
SQ Sequence 1320 AA;
Query Match 92.2%; Score 5370.2; DB 7; Length 1320;
Best Local Similarity 92.3%; Pred. No. 3e-146;
Matches 1013; Conservative 8; Mismatches 28; Indels 48; Gaps 2;
QY 1 MAWKTLPIYLLLSVFVIQQVSSQDLSSCAGRGEGYSRDATCNDYHCQHYECCPDF 60
Db 1 MAWKTLPIYLLLSVFVIQQVSSQ-----
QY 61 KRVTALSCKGRCFESFERGECDDAQCCKYDKCCPDYSEFCABEHSYSENCSSSS 120
Db -----ELSCKGRCFESFERGECDDAQCCKYDKCCPDYSEFCABEHSYSENCSSSS 120
QY 26 -----ELSCKGRCFESFERGECDDAQCCKYDKCCPDYSEFCABEHSYSENCSSSS 120
Db -----ELSCKGRCFESFERGECDDAQCCKYDKCCPDYSEFCABEHSYSENCSSSS 120
QY 121 SSSSSSSSTIWKIKSSKNSAANRELQKL-----KVKDNKKKRTKKKPTPKPPVDEAG 173
Db -----KVKDNKKKRTKKKPTPKPPVDEAG 173
QY 80 PPSGASQTIKSTTKRSPKPKPKKKTKKVKIESEEITEVKNKNKRTKKKPTPKPPVDEAG 139
Db -----KVKDNKKKRTKKKPTPKPPVDEAG 139
QY 174 SGLDNGDFKVTTPDTSTTOHKNKYSTSPKITTAKPINRPSLPNSDTSKTSLTVNKETT 233
Db SGLDNGDFKVTTPDTSTTOHKNKYSTSPKITTAKPINRPSLPNSDTSKTSLTVNKETT 199

[illegible]

RESULT 10

ADK67911

ID ADK67911 standard; protein; 1270 AA.

XX AC ADK67911:

XX
ADK61911;

DT 06-MAY-2004 (first entry)

```

XX
DE Human cytomegalovirus (HCMV)  ncj:montido

```

antiasthmatic; antiinflammatory; antidiabetic; neuroprotective; muscular-gen.; antiarthritic; osteopathic; hepatotropic; antipsoriatic; virucide; fungicide; antiparasitic; protozoacide; antihelminthic; cytostatic; gene therapy.

Homo sapiens.

| Key | Location/Qualifiers |
|-----|---------------------|
|-----|---------------------|

Peptide 1.1.29

```

reporter
i: 29
/label= Signal_peptide
/note= "Spans residues 1 to 18, 20, 21, 24, 27 or 29
according to identification method"

```

WO2004013292-A2.

12-FEB-2004

30-JUL-2003: 2003WO-IIS024084

02-AUG-2002: 2002UIS-0400810P

02-AUG-2002; 2002US-0408TOP.
19-SEP-2002; 2002US-0412197P

19-SEP-2002; 2002US-0412197P.
04-OCT-2002: 2002US-0416004P

04-OCT-2002; 2002US-0416004P.
08-NOV-2002; 2002US-0424862P

(TN0V-1) TN0VTE C00D

Elliot & Weiss
Whaver &
Eaton &
Sweeney &
Maxwell & J.D.

Elliot VS, Khare R, Tran UK, Swarnakar A, Marquis JP; Pichardson TW, Emerling BM, Lindquist EA, Chawla NK, Ramkumar J.

Richard E. ...

0004 1 2 5 1 9 0 / 1 5

WPI; 2004-157116/

New extracellular messengers and nucleic acids, useful for diagnosing, treating or preventing e.g. allergies, anemia, Crohn's disease, diabetes mellitus, myasthenia gravis, osteoporosis, psoriasis, cancers or autoimmune thyroiditis.

Claim 59: SEQ ID NO 4: 165bp: English.

The present sequence is that of novel human extracellular messenger (EXMES) Incyte ID No: 7513017CD1 polypeptide. The protein shows homology to human megakaryocyte stimulating factor. The invention provides EXMES to human megakaryocytes and polypeptides, as well as expression vectors, host cells, antibodies, agonists and antagonists, and methods for diagnosing, treating or preventing disorders associated with aberrant expression of EXMES, especially autoimmune and inflammatory disorders, cell proliferative disorders and endocrine disorders, e.g. adult respiratory distress syndrome, allergy, anaemia, erythroblastosis foetalis, Crohn's disease, diabetes mellitus, myasthenia gravis, osteoarthritis, fungal, osteoporosis, pancreatitis, psoriasis, viral, bacterial, fungal, parasitic, protozoal or helminthic infections, cancers, autoimmune thyroiditis, cretinism, Plummer's disease or thyroid carcinoma. Embodiments also provide methods for using the purified EXMES and/or their encoding polynucleotides for facilitating the drug discovery process, including determining of efficacy, dosage, toxicity and pharmacology, and for investigating the pathogenesis of diseases and medical conditions.

Sequence 1270 AA;

| | | | | | |
|--------|-------|---------|-------|--------|-------|
| 91.7%: | Score | 5344.6: | DB 8: | Length | 1270: |
|--------|-------|---------|-------|--------|-------|

91.7%; SCORE 3347.8; DB 8
92.38%; Pred. No. 1.6e-145; Pred. No. 1.6e-145;

| Seq. ID | Accession | Length | Similarity | Pos. NO. | Indels | Mismatches | Gaps |
|---------|-----------|--------|------------|----------|--------|------------|------|
| 1 | U00001.1 | 1089 | 100% | 0 | 0 | 0 | 0 |
| 2 | U00002.1 | 1089 | 100% | 0 | 0 | 0 | 0 |
| 3 | U00003.1 | 1089 | 100% | 0 | 0 | 0 | 0 |
| 4 | U00004.1 | 1089 | 100% | 0 | 0 | 0 | 0 |
| 5 | U00005.1 | 1089 | 100% | 0 | 0 | 0 | 0 |
| 6 | U00006.1 | 1089 | 100% | 0 | 0 | 0 | 0 |
| 7 | U00007.1 | 1089 | 100% | 0 | 0 | 0 | 0 |
| 8 | U00008.1 | 1089 | 100% | 0 | 0 | 0 | 0 |
| 9 | U00009.1 | 1089 | 100% | 0 | 0 | 0 | 0 |
| 10 | U00010.1 | 1089 | 100% | 0 | 0 | 0 | 0 |
| 11 | U00011.1 | 1089 | 100% | 0 | 0 | 0 | 0 |
| 12 | U00012.1 | 1089 | 100% | 0 | 0 | 0 | 0 |
| 13 | U00013.1 | 1089 | 100% | 0 | 0 | 0 | 0 |
| 14 | U00014.1 | 1089 | 100% | 0 | 0 | 0 | 0 |
| 15 | U00015.1 | 1089 | 100% | 0 | 0 | 0 | 0 |
| 16 | U00016.1 | 1089 | 100% | 0 | 0 | 0 | 0 |
| 17 | U00017.1 | 1089 | 100% | 0 | 0 | 0 | 0 |
| 18 | U00018.1 | 1089 | 100% | 0 | 0 | 0 | 0 |
| 19 | U00019.1 | 1089 | 100% | 0 | 0 | 0 | 0 |
| 20 | U00020.1 | 1089 | 100% | 0 | 0 | 0 | 0 |
| 21 | U00021.1 | 1089 | 100% | 0 | 0 | 0 | 0 |
| 22 | U00022.1 | 1089 | 100% | 0 | 0 | 0 | 0 |
| 23 | U00023.1 | 1089 | 100% | 0 | 0 | 0 | 0 |
| 24 | U00024.1 | 1089 | 100% | 0 | 0 | 0 | 0 |
| 25 | U00025.1 | 1089 | 100% | 0 | 0 | 0 | 0 |
| 26 | U00026.1 | 1089 | 100% | 0 | 0 | 0 | 0 |
| 27 | U00027.1 | 1089 | 100% | 0 | 0 | 0 | 0 |
| 28 | U00028.1 | 1089 | 100% | 0 | 0 | 0 | 0 |
| 29 | U00029.1 | 1089 | 100% | 0 | 0 | 0 | 0 |
| 30 | U00030.1 | 1089 | 100% | 0 | 0 | 0 | 0 |
| 31 | U00031.1 | 1089 | 100% | 0 | 0 | 0 | 0 |
| 32 | U00032.1 | 1089 | 100% | 0 | 0 | 0 | 0 |
| 33 | U00033.1 | 1089 | 100% | 0 | 0 | 0 | 0 |
| 34 | U00034.1 | 1089 | 100% | 0 | 0 | 0 | 0 |
| 35 | U00035.1 | 1089 | 100% | 0 | 0 | 0 | 0 |
| 36 | U00036.1 | 1089 | 100% | 0 | 0 | 0 | 0 |
| 37 | U00037.1 | 1089 | 100% | 0 | 0 | 0 | 0 |
| 38 | U00038.1 | 1089 | 100% | 0 | 0 | 0 | 0 |
| 39 | U00039.1 | 1089 | 100% | 0 | 0 | 0 | 0 |
| 40 | U00040.1 | 1089 | 100% | 0 | 0 | 0 | 0 |
| 41 | U00041.1 | 1089 | 100% | 0 | 0 | 0 | 0 |
| 42 | U00042.1 | 1089 | 100% | 0 | 0 | 0 | 0 |
| 43 | U00043.1 | 1089 | 100% | 0 | 0 | 0 | 0 |
| 44 | U00044.1 | 1089 | 100% | 0 | 0 | 0 | 0 |
| 45 | U00045.1 | 1089 | 100% | 0 | 0 | 0 | 0 |
| 46 | U00046.1 | 1089 | 100% | 0 | 0 | 0 | 0 |
| 47 | U00047.1 | 1089 | 100% | 0 | 0 | 0 | 0 |
| 48 | U00048.1 | 1089 | 100% | 0 | 0 | 0 | 0 |
| 49 | U00049.1 | 1089 | 100% | 0 | 0 | 0 | 0 |
| 50 | U00050.1 | 1089 | 100% | 0 | 0 | 0 | 0 |
| 51 | U00051.1 | 1089 | 100% | 0 | 0 | 0 | 0 |
| 52 | U00052.1 | 1089 | 100% | 0 | 0 | 0 | 0 |
| 53 | U00053.1 | 1089 | 100% | 0 | 0 | 0 | |

1 MAWKTLPIYLLLLLSVFVIQQVSSQDLSSCAGRCGEGYSRDATCNCNDYNCQHYMECCPDF 60

1 MAWKTLPIYLLLLLSVFVIQQVSSQ-----25

61 KRVCTAELSCGRCFESFERGRECDCAOCKKYDKCCPDYESFCAEEHVSSENQESSSS 120

26 -----ELCKGRCFESFERGRECDCAOCKYDKCCPDYESFCAE----- 65


```

Db 361 ETPTPTTSEVSTTTTKEPTTIHKSPDESTPELSAEPPTKALENSPKBEGVETTKTTPAAT 420
QY 858 KPEMTTAKDKITERDLRTTPEPTTAAAPKVTKETATTTTEKTTESKITATTTQVTSITTD 917
Db 421 KPEMTTAKDKITERDLRTTPEPTTAAAPKVTKETATTTTEKTTESKITATTTQVTSITTD 480
QY 918 TTPFKITLTKTTLAPKVTTKKTIITTEIMNKPBEETAKPKDRATNSKATTPKPKPTKA 977
Db 481 TTPFKITLTKTTLAPKVTTKKTIITTEIMNKPBEETAKPKDRATNSKATTPKPKPTKA 540
QY 978 PKKPTS 983
Db 541 PKKPTS 546

RESULT 12
ABU53253
ID ABU53253 standard; protein; 551 AA.
XX
AC ABU53253;
XX
DT 14-APR-2003 (first entry)
XX
DE Human testes-derived DKFzpthes3_4019 homologue #2.
XX
KW Human; gene therapy; vaccine; disease treatment; detection.
XX
OS Homo sapiens.
XX
PN WO200112659-A2.
XX
PD 22-FEB-2001.
XX
PF 18-AUG-2000; 2000WO-IB001496.
XX
PR 18-AUG-1999; 99US-0149499P.
PR 28-SEP-1999; 99US-0156503P.
XX
PA (GEHU-) GERMAN HUMAN GENOME PROJECT.
XX
PI Wiemann S;
XX
DR WPI; 2001-327840/34.
XX
PT Nucleic acids having the sequences of clones isolated from libraries of
PT different human tissues, useful in recombinant DNA methodologies.
XX
PS Example III; Page 892-893; 1095pp; English.
XX
CC This invention describes novel polynucleotides and polypeptides isolated
CC from human cDNA libraries which can be used for gene therapy or in
CC vaccines. The polynucleotides of the invention and antibodies encoded by
CC them may be used in the prevention, diagnosis and treatment of diseases
CC associated with inappropriate polypeptide expression. The products of the
CC invention may also be used to identify modulators of expression and
CC activity and to down regulate expression and activity. The antibodies of
CC the invention may also be used as diagnostic agents for detecting the
CC presence of polypeptides in samples. This sequence represents a homologue
CC of a polypeptide described in the disclosure of the invention
XX
SQ Sequence 551 AA;

Query Match 49.6%; Score 2888; DB 4; Length 551;
Best Local Similarity 100.0%; Pred. No. 2.4e-75;
Matches 544; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 547 TKKPAPTAPKEPAPTTPKETAPTTPKLTPTTPEKLAPTTPKEPAPTTPPELAPTTPPEP 606
Db 1 TKKPAPTAPKEPAPTTPKETAPTTPKLTPTTPEKLAPTTPKEPAPTTPPELAPTTPPEP 60
QY 607 TPTTPEPAPTTPKAAAPNTKPEPAPTTPKEPAPTTPKEPAPTTPKETAPTTPKGTAPT 666
Db 61 TPTTPEPAPTTPKAAAPNTKPEPAPTTPKEPAPTTPKEPAPTTPKETAPTTPKGTAPT 120

```

```

QY 667 LKSPAPTTPKKPAPKELAPTTPKPTSTTSKDPAPTTPKGTAPTTPKEPAPTTPKEPAPT 726
Db 121 LKSPAPTTPKKPAPKELAPTTPKPTSTTSKDPAPTTPKGTAPTTPKEPAPTTPKEPAPT 180
QY 727 TPKGTAPTTLKBPAPTTPKKPAPKELAPTTPKGTSTTSKDPAPTTPKGTAPTTPKEPAPT 786
Db 181 TPKGTAPTTLKBPAPTTPKKPAPKELAPTTPKGTSTTSKDPAPTTPKGTAPTTPKEPAPT 240
QY 787 TTPKKPAPTTPPEPTTSEVSTPTTTPKPTTIHKSPDESTPELSAEPPTKALENSPKBP 846
Db 241 TTPKKPAPTTPPEPTTSEVSTPTTTPKPTTIHKSPDESTPELSAEPPTKALENSPKBP 300
QY 847 GVPTTKTPAATKPEMTTAKDKITERDLRTTPEPTTAAAPKVTKETATTTTEKTTESKITAT 906
Db 301 GVPTTKTPAATKPEMTTAKDKITERDLRTTPEPTTAAAPKVTKETATTTTEKTTESKITAT 360
QY 907 TTVTSITTTQDITTPFKITLTKTTLAPKVTTKKTIITTEIMNKPBEETAKPKDRATNSKA 966
Db 361 TTVTSITTTQDITTPFKITLTKTTLAPKVTTKKTIITTEIMNKPBEETAKPKDRATNSKA 420
QY 967 TTPKPQKPTKAPKKPTSTTKPKTTPRVRKPKTTPRKMVSTMPKLNPTSRIAEAMLQTT 1026
Db 421 TTPKPQKPTKAPKKPTSTTKPKTTPRVRKPKTTPRKMVSTMPKLNPTSRIAEAMLQTT 480
QY 1027 TRNQTPNSKLVEVNPKSEDAAGBGETPHMLLRPHVFMPEVTPDMDYLPVFNQGIIN 1086
Db 481 TRNQTPNSKLVEVNPKSEDAAGBGETPHMLLRPHVFMPEVTPDMDYLPVFNQGIIN 540
QY 1087 PMLS 1090
Db 541 PMLS 544

RESULT 13
AAB29778
ID AAB29778 standard; protein; 902 AA.
XX
AC AAB29778;
XX
DT 28-FEB-2001 (first entry)
XX
DE Human MSF-derived tribonectin.
XX
KW Human tribonectin; MSF; megakaryocyte stimulating factor;
KW alternative splicing; joint boundary lubricant; O-linked oligosaccharide;
KW osteoarthritis; tribosupplementation; tissue adhesion inhibition;
KW friction coefficient reduction; gene therapy; antiarthritic; osteopathic.
XX
OS Homo sapiens.
XX
PN WO200064930-A2.
XX
DD 02-NOV-2000.
XX
PF 24-APR-2000; 2000WO-US010953.
XX
PR 23-APR-1999; 99US-00298970.
XX
PA (RHOD-) RHODE ISLAND HOSPITAL LIFESPAN PARTNER.
XX
PI Jay GD;
XX
DR WPI; 2001-024673/03.
XX
PT Novel tribonectin polypeptide useful as lubricant for treating
PT osteoarthritis, comprises O-linked lubricating moiety.
XX
PS Disclosure; Fig 1; 47pp; English.
XX
CC The invention relates to a human tribonectin which is a product of
CC alternative splicing of the human MSF (megakaryocyte stimulating factor)
CC gene. The tribonectin has at least one O-linked oligosaccharide

```

CC lubricating moiety and has a polypeptide sequence comprising 1-76 repeats
CC of a motif having at least 50% identity to the sequence KEPAPTT
CC (AAB29774). The invention also relates to a nucleic acid encoding a human
CC MSP-derived tribonectin; a biocompatible composition comprising a human
CC tribonectin for inhibiting tissue adhesion formation; and a method of
CC diagnosing osteoarthritis or a predisposition to osteoarthritis by
CC measuring the amount of MSP or its fragment in a biological sample of a
CC mammal, wherein an increased amount of MSP compared to a control
CC indicates the presence of or predisposition to developing osteoarthritis.
CC The tribonectin and DNA encoding it are useful in the treatment of
CC osteoarthritis where they may be used for lubricating mammalian joints,
CC such as articulating joints of humans, dogs or horses. The tribonectin,
CC when formulated as a membrane, foam, gel or fibre, is useful for
CC inhibiting adhesion between two surfaces such as the injured tissues of a
CC mammal, where the injury is caused by a surgical insertion or trauma, or
CC an artificial device e.g., an orthopaedic implant. In particular, one of
CC the surfaces is pericardial tissue. DNA encoding a tribonectin may be
CC used in gene therapy. The present sequence represents a substantial
CC portion of a human MSP-derived tribonectin
XX
XX Sequence 902 AA;
Query Match 49.4%; Score 2880.1; DB 4; Length 902;
Best Local Similarity 71.4%; Pred. No. 8.1e-75;
Matches 675; Conservative 27; Mismatches 74; Indels 169; Gaps 66;
QY 150 VKDNKNRTKKKPTKPPVVDVDEAGSLDNGDFKVTTPDSTTQHNNKYSTSPKITTAKPIN 209
Db 1 VKDNKNRTKKKPTKPPVVDVDEAGSLDNGDFKVTTPDSTTQHNNKYSTSPKITTAKPIN 60
QY 210 PRSLPNSDTSKETSLSLVNKEITVETKTTTNTKOTSDGKEKTTSAKETOSIEKTSNAK 269
Db 61 PRSLPNSDTSKETSLSLVNKEITVETKTTTNTKOTSDGKEKTTSAKETOSIEKTSNAK 120
QY 270 DLAPTSKVLAKPTPKAETTTKGPALTTPKEPTTTPKEPASTTPKEPTTTIKSAPTPPK 329
Db 121 DLAPTSKVLAKPTPKAETTTKGPAL-----TTPKEPASTTPKEPTTTIKSAPTPPK 172
QY 330 EPAPTTKSNAPTPKEPAPTTTKEPAPTTKBPAPTTTKEPAPTTTKSAPTPPKBPAPTT 389
Db 173 EPAPTTKSNAPTPKEPAPTTTKEPAPTTKBPAPTTTKEPAPTTTKEPAPTTTKEPAPTT 228
QY 390 PKKAPPTTKEPAPTTPKBPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAP 449
Db 229 -KEPAPTT-KEPAPTT-KEPAPTT-KEPAPTT-KEPAPTT-KEPAPTT-KEPAPTT-KEPAP 281
QY 450 PTPPKBPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKE 509
Db 282 PTT-KEPAP-TTKEPAPTT-KEPAPTTKGPAP-TTKEPAPTTKEPAPTT-KEPAP-TTKE 335
QY 510 PAPTTKEPAPTTPKBPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTT 569
Db 336 PAPTT-KEPAPTT-KEPAPTT-KEPAPTT-KEPAPTT-KEPAPTT-KEPAPTT-KEPAPTT 388
QY 570 TPKKLTPTTPEKLAPTTPKBPAPTTTPEELAPTTPEPTTTPPEPAPTTTPEKAAAPNTPKE 629
Db 389 T-KEPAPTTKEP-APTKE-PAPTT-KEPAPTT-KEPAPTT-KEPAPTT-KEPAPTT-KE 440
QY 630 PAPTTKEPAPTTPKBPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTT 687
Db 441 PAPTT-KEPAPTT-KEPAPTT-KEPAPTT-KEPAPTT-KEPAPTT-KEPAPTT-KEPAPTT 493
QY 688 TKEPTSTSDKAPPTPKGTAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTT 747
Db 494 TKEPAPTTKE-PAPTT-KEPAPTT-KEPAPTT-KEPAPTT-KEPAPTT-KEPAPTT-KEP 546
QY 748 AP---KELAPTTTKGPTSTSDKAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKE 805
Db 547 APTTKEPAP-TTKEPAPTTTKE-PAPTT-KEPAPTT-KEPAPTT-KEPAPTTTKE-PAPTTK 600
QY 806 EVSTPTTTKEPTTIHKSPPDSTSELSABPTPKALENSPKBPVGVTPTTKTAANKPEWTTTA 865
Db 601 EPA--PTTKEPAPTTKEPAPTT---KEPAP-----TTKEP-APTKEPAPTT----- 640

QY 866 KDKTTERDLRTTPETTTAAAPKMTKETATTTEKTTESKITATTTOVISTTTODTTPFKITT 925
Db 641 -----TKEPAPTT-----KEPAPTTKEPAPTTKEPAP----- 667
QY 926 LKTTTLAPKVTTKKTTTTEIMNKPBETAKPKDRATNSKATTPKPKOKPKAPKPKPTSTK 985
Db 668 -----TTKEPAPTTTKEP-APT-TK 684
QY 986 PKTMTPRVRKPKTTPTRKMTSTMPELNPTSPRIAEAMLQTTTRNQTPNSKLVENVPKSE 1045
Db 685 EP-----APTTPTRKMTSTMPELNPTSPRIAEAML-TTTRNQTPNSKLVENVPKSE 735
QY 1046 DAGAEGETPHMLLRPHVFPVETPDMYLPVFNQGIINPMLS 1090
Db 736 DAGAEGETPHMLLRPHVFPVETPDMYLPVFNQGIINPMLS 780
RESULT 14
ABUS3254
ID ABUS3254 standard; protein; 513 AA.
XX AC ABUS3254;
XX DT 14-APR-2003 (first entry)
XX DE Human testes-derived DRFZphes3_4019 homologue #3.
XX KW Human; gene therapy; vaccine; disease treatment; detection.
XX OS Homo sapiens.
XX PN WO200112659-A2.
XX PD 22-FEB-2001.
XX PF 18-AUG-2000; 2000WO-IB001496.
XX PR 18-AUG-1999; 99US-0149499P.
XX PR 28-SEP-1999; 99US-0156503P.
XX PA (GEHU-) GERMAN HUMAN GENOME PROJECT.
XX PI Wiemann S;
XX DR WPI; 2001-327840/34.
XX PT Nucleic acids having the sequences of clones isolated from libraries of
XX different human tissues, useful in recombinant DNA methodologies.
XX PS Example III; Page 893; 1095pp; English.
XX CC This invention describes novel polynucleotides and polypeptides isolated
XX from human cDNA libraries which can be used for gene therapy or in
XX vaccines. The polynucleotides of the invention and antibodies encoded by
XX them may be used in the prevention, diagnosis and treatment of diseases
XX associated with inappropriate polypeptide expression. The products of the
XX invention may also be used to identify modulators of expression and
XX activity and to down regulate expression and activity. The antibodies of
XX the invention may also be used as diagnostic agents for detecting the
XX presence of polypeptides in samples. This sequence represents a homologue
XX of a polypeptide described in the disclosure of the invention
SQ Sequence 513 AA;
Query Match 47.3%; Score 2757; DB 4; Length 513;
Best Local Similarity 100.0%; Pred. No. 1.3e-71;
Matches 513; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 157 RTKKKPTPKPPVDEAGSLDNGDFKVTTPDSTTQHNNKYSTSPKITTAKPINRPSLPP 216
Db 1 RTKKKPTPKPPVDEAGSLDNGDFKVTTPDSTTQHNNKYSTSPKITTAKPINRPSLPP 60


```

QY 217 NSDTSKETSITVNNKETTIVETKTTTNNKQSTSDGKEKTSKETSIAKTSKADLAPTSK 276
DB 61 NSDTSKETSITVNNKETTIVETKTTTNNKQSTSDGKEKTSKETSIAKTSKADLAPTSK 120
QY 277 VLAKEPTPKAEITTKGPAITTPKEPTTPPKEPASTTPKEPTTTIKSAPTTPKEPAPTTT 336
DB 121 VLAKEPTPKAEITTKGPAITTPKEPTTPPKEPASTTPKEPTTTIKSAPTTPKEPAPTTT 180
QY 337 KSAPITPKEPAPTTTKGPAITTPKEPTTPPKEPASTTPKEPTTTIKSAPTTPKEPAPTT 396
DB 181 KSAPITPKEPAPTTTKGPAITTPKEPTTPPKEPASTTPKEPTTTIKSAPTTPKEPAPTT 240
QY 397 TPKEPAPTTPKEPAPTTTKGPAITTPKEPTTPPKEPASTTPKEPTTTIKSAPTTPKEPAPTT 456
DB 241 TPKEPAPTTPKEPAPTTTKGPAITTPKEPTTPPKEPASTTPKEPTTTIKSAPTTPKEPAPTT 300
QY 457 APTTTKEPSPTTPKEPAPTTTKGPAITTPKEPTTPPKEPASTTPKEPTTTIKSAPTTPKEPAPTT 516
DB 301 APTTTKEPSPTTPKEPAPTTTKGPAITTPKEPTTPPKEPASTTPKEPTTTIKSAPTTPKEPAPTT 360
QY 517 EPAPITPKKPAITPKEPAPTTTKGPAITTPKEPTTPPKEPASTTPKEPTTTIKSAPTTPKEPAPTT 576
DB 361 EPAPITPKKPAITPKEPAPTTTKGPAITTPKEPTTPPKEPASTTPKEPTTTIKSAPTTPKEPAPTT 420
QY 577 TPPEKLAITTPKEPAPTTTPPEELAPTTPEEPTTPPEEAPTTTPKAAAPNTPKEPAPTTPK 636
DB 421 TPPEKLAITTPKEPAPTTTPPEELAPTTPEEPTTPPEEAPTTTPKAAAPNTPKEPAPTTPK 480
QY 637 EPAPITPKKPAITPKEPAPTTTKGPAITTPKEPTTPPKEPASTTPKEPTTTIKSAPTTPKEPAPTT 669
DB 481 EPAPITPKKPAITPKEPAPTTTKGPAITTPKEPTTPPKEPASTTPKEPTTTIKSAPTTPKEPAPTT 513

RESULT 15
ID AAR80041 standard; protein; 452 AA.
XX
AC AAR80041;
XX
DT 25-MAR-2003 (revised)
DT 10-APR-1996 (first entry)
XX
DE Human megakaryocytopoietin protein.
XX
KW Human; megakaryocytopoietin; wheat germ agglutinin; heparin;
KW megakaryocyte; aplastic anaemia; thrombocytopenia; hematocytopenia;
KW multipotential stem cell.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT Misc-difference 393..396
FT /note= "unspecified amino acids"
FT Misc-difference 444..446
FT /note= "unspecified amino acids"
XX
PN WO9523861-Al.
XX
PD 08-SEP-1995.
XX
PF 06-MAR-1995; 95WO-CN000015.
XX
PR 04-MAR-1994; 94CN-00112066.
XX
PA (SHAN-) SHANGHAI BEITE BIOTECHNOLOGY CO LTD.
XX
PI Gu X, Han Z, Shen Q;
XX
DR WPI; 1995-320576/41.
DR N-PSDB; AAT04546.
XX
PT New haematopoietic cell growth factor - used for treating
PT thrombocytopenia and hematocytopenia.

```

```

XX
PS
XX
CC This sequence represents the human megakaryocytopoietin (MPO) protein.
CC This sequence was purified using a carrier which can couple wheat germ
CC agglutinin and heparin to separate MPO. Fragments of this sequence (see
CC AAR80039 and AAR80040) were used to produce the amplification primers
CC shown in AAT04544 and AAT04545. The fragments amplified by these primers
CC can then be used as probes to screen human cDNA libraries for MPO cDNA.
CC The MPO cDNA can then be inserted into a plasmid which is used to
CC transform cells to produce MPO. The MPO sequence is capable of promoting
CC colony formation of megakaryocytes, enlarging the size of megakaryocytes
CC and stimulating the proliferation of multipotential stem cells. The
CC factor may be used for treating thrombocytopenia and hematocytopenia. The
CC purification method can be used to isolate MPO from human urine or serum
CC of patients with aplastic anaemia, and from animal blood or urine by
CC radiation exposing the animals to induce aplastic anaemia. (Updated on 25
CC -MAR-2003 to correct PA field.)
XX
SQ Sequence 452 AA;
Query Match 34.2%; Score 1994.2; DB 2; Length 452;
Best Local Similarity 83.3%; Pred. No. 9.7e-50;
Matches 378; Conservative 11; Mismatches 47; Indels 18; Gaps 2;
QY 1 MAWKTLPIVILLLLSVFVIQQVSSQDLSSCAGCGEGYSRDATCNCYNQCHYMECCPDF 60
DB 1 MAWKTLPIVILLLLSVFVIQQVSSQDLSSCAGCGEGYSRDATCNCYNQCHYMECCPDF 60
QY 61 KRVTAEELCKGRCFESFERGECDCDAQCKYDKCCPDYSEFCABEHSVSENQSSSSS 120
DB 61 KRVTAEELCKGRCFESFERGECDCDAQCKYDKCCPDYSEFCABEHSVSENQSSSSS 120
QY 121 SSSSSSTIWKIKSSXNSAANRELQKL-----KVKDNKNRKKKPTPKPPVVDEAG 173
DB 121 PPSGASQTIKSTTKRSFKNKKYKVVIESEITEVDKNKNRKKKPTPKPPVVDEAG 180
QY 174 SGLDNGDFKVTTPDSTTTOHNVKSTSPKITTAKPINRPSLPNSDTSKETSITVNNKETT 233
DB 181 SGLDNGDFKVTTPDSTTTOHNVKSTSPKITTAKPINRPSLPNSDTSKETSITVNNKETT 240
QY 234 VETKETTTNKQSTSDGKEKTSKETSIAKTSKADLAPTSKVLAKTPKAEITTKGPA 293
DB 241 VETKETTTNKQSTSDGKEKTSKETSIAKTSKADLAPTSKVLAKTPKAEITTKGPA 300
QY 294 LPTPKEPPTTPKEPASTTPKEPTTTIKSAPTTPKEPAPTTTKSAPTTPKEPAPTTTKE 353
DB 301 LPTPKEPPTTPKEPASTTPKEPTTTIKSAPTTPKEPAPTTTKSAPTTPKEPAPTTTKE 360
QY 354 PAPTTPKEPAPTTTKGPAITTPKEPTTPPKEPASTTPKEPTTTIKSAPTTPKEPAPTTT 413
DB 361 PAPTTPKEPAPTTTKGPAITTPKEPTTPPKEPASTTPKEPTTTIKSAPTTPKEPAPTTT 413
QY 414 PKEPAPTTPKEPAPTTPKEPAPTTAPKPAITPKKE 447
DB 410 PRSLHPTPKEPAPTTPKEPAPTTAPKPAITPKKE 443

Search completed: October 13, 2004, 11:36:46
Job time : 93.8789 secs

```


GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: October 13, 2004, 11:23:49 ; Search time 18.9474 Seconds
(without alignments)
3815.116 Million cell updates/sec

Title: SEQ1-B
Perfect score: 5826
Sequence: 1 MANKTPIVILLLSVFVIQ.....DMDYLPVNPQGIINPMLS 1090

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.1

Searched: 478139 seqs, 66318000 residues

Total number of hits satisfying chosen parameters: 478139

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued Patents AA:*
1: /cgn2_6/ptodata/1/iaa/5A-COMB.pep.*
2: /cgn2_6/ptodata/1/iaa/5B-COMB.pep.*
3: /cgn2_6/ptodata/1/iaa/6A-COMB.pep.*
4: /cgn2_6/ptodata/1/iaa/6B-COMB.pep.*
5: /cgn2_6/ptodata/1/iaa/PTUS-COMB.pep.*
6: /cgn2_6/ptodata/1/iaa/backfiles1.pep.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | ID | Description |
|------------|--------|-------------|--------|----------------------|--------------------|
| 1 | 5826 | 100.0 | 1354 | 4 US-07-757-022B-48 | Sequence 48, Appl |
| 2 | 5811 | 99.7 | 1140 | 4 US-07-757-022B-104 | Sequence 104, Appl |
| 3 | 5811 | 99.7 | 1404 | 4 US-07-757-022B-2 | Sequence 2, Appl |
| 4 | 5811 | 99.7 | 1404 | 4 US-07-757-022B-62 | Sequence 62, Appl |
| 5 | 5811 | 99.7 | 1404 | 4 US-09-298-970A-1 | Sequence 1, Appl |
| 6 | 5788 | 99.3 | 1404 | 4 US-10-164-595-78 | Sequence 78, Appl |
| 7 | 5643.3 | 96.9 | 1361 | 4 US-07-757-022B-40 | Sequence 40, Appl |
| 8 | 5608.7 | 96.3 | 1311 | 4 US-07-757-022B-42 | Sequence 42, Appl |
| 9 | 5571 | 95.6 | 1314 | 4 US-07-757-022B-50 | Sequence 50, Appl |
| 10 | 5561.9 | 95.5 | 1049 | 4 US-07-757-022B-58 | Sequence 58, Appl |
| 11 | 5561.9 | 95.5 | 1313 | 4 US-07-757-022B-142 | Sequence 142, Appl |
| 12 | 5546.9 | 95.2 | 1363 | 4 US-07-757-022B-52 | Sequence 52, Appl |
| 13 | 5486.7 | 93.8 | 1022 | 4 US-07-757-022B-84 | Sequence 84, Appl |
| 14 | 5379.2 | 92.3 | 1320 | 4 US-07-757-022B-46 | Sequence 46, Appl |
| 15 | 5379.2 | 92.3 | 1320 | 4 US-07-757-022B-60 | Sequence 60, Appl |
| 16 | 5370.2 | 92.2 | 1320 | 4 US-10-164-595-58 | Sequence 58, Appl |
| 17 | 5344.6 | 91.7 | 1038 | 4 US-07-757-022B-74 | Sequence 74, Appl |
| 18 | 5344.6 | 91.7 | 1270 | 4 US-07-757-022B-44 | Sequence 44, Appl |
| 19 | 5011 | 86.0 | 941 | 4 US-07-757-022B-14 | Sequence 14, Appl |
| 20 | 1394.5 | 23.9 | 5179 | 4 US-09-538-092-1258 | Sequence 1258, Ap |
| 21 | 1091.9 | 18.7 | 8991 | 4 US-08-714-741-52 | Sequence 32, Appl |
| 22 | 914 | 15.7 | 220 | 4 US-07-757-022B-96 | Sequence 96, Appl |
| 23 | 904.3 | 15.5 | 207 | 4 US-07-757-022B-116 | Sequence 116, Appl |
| 24 | 904.3 | 15.5 | 207 | 4 US-07-757-022B-136 | Sequence 136, Appl |
| 25 | 869.7 | 14.9 | 157 | 4 US-07-757-022B-102 | Sequence 102, Appl |
| 26 | 869.7 | 14.9 | 157 | 4 US-07-757-022B-114 | Sequence 114, Appl |
| 27 | 852 | 14.6 | 209 | 4 US-07-757-022B-94 | Sequence 94, Appl |

Sequence 132, Appl
Sequence 54, Appl
Sequence 98, Appl
Sequence 22, Appl
Sequence 21, Appl
Sequence 92, Appl
Sequence 1, Appl
Sequence 2, Appl
Sequence 70, Appl
Sequence 30, Appl
Sequence 90, Appl
Sequence 5, Appl
Sequence 5, Appl
Sequence 1142, Ap
Sequence 140, Appl
Sequence 5, Appl
Sequence 6, Appl
Sequence 6, Appl

28 847 14.5 208 4 US-07-757-022B-132
29 838.2 14.4 463 4 US-07-757-022B-54
30 831.7 14.3 3256 4 US-09-919-172-98
31 831.7 14.3 3256 4 US-09-919-172-98
32 831.7 14.3 3256 4 US-09-919-172-98
33 826 14.2 204 4 US-07-757-022B-92
34 820.9 14.1 3118 3 US-09-579-181-1
35 813.6 14.0 2972 3 US-09-579-181-2
36 813.2 14.0 296 4 US-07-757-022B-70
37 801.6 13.8 231 4 US-07-757-022B-30
38 767 13.2 192 4 US-07-757-022B-90
39 764.4 13.1 1837 3 US-08-928-361B-5
40 764.4 13.1 1837 3 US-08-928-361B-5
41 761.8 13.1 2142 4 US-09-588-995A-5
42 748.7 12.9 132 4 US-07-757-022B-140
43 745.9 12.8 1721 3 US-08-700-551-5
44 745.9 12.8 1721 3 US-08-928-361B-6
45 745.9 12.8 1721 4 US-09-588-995A-6

ALIGNMENTS

RESULT 1
US-07-757-022B-48
; Sequence 48, Application US/07757022B
; Patent No. 6433142
; GENERAL INFORMATION:
; APPLICANT: Gesner, Thomas G.
; APPLICANT: Clark, Stephen C.
; APPLICANT: Turner, Katherine
; APPLICANT: Hewick, Rodney M.
; TITLE OF INVENTION: Megakaryocyte Stimulating Factors
; NUMBER OF SEQUENCES: 143
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genetics Institute, Inc.
; STREET: 87 CambridgePark Drive
; CITY: Cambridge
; STATE: Massachusetts
; COUNTRY: U.S.A.
; ZIP: 02140
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07757.022B
; FILING DATE: 19910910
; CLASSIFICATION: 530
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/643,502
; FILING DATE: 18-JAN-1991
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/546,114
; FILING DATE: 29-JUN-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/457,196
; FILING DATE: 29-DEC-1989
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/390,901
; FILING DATE: 08-AUG-1989
; ATTORNEY/AGENT INFORMATION:
; NAME: Coert, Luann
; REGISTRATION NUMBER: 31,822
; REFERENCE/DOCKET NUMBER: GI 5190
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617)876-1170
; TELEFAX: (617)876-5851
; INFORMATION FOR SEQ ID NO: 48:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1354 amino acids
; TYPE: AMINO ACID

TOPOLOGY: linear
MOLECULE TYPE: protein
US-07-757-022B-48

Query Match 100.0%; Score 5826; DB 4; Length 1354;
Best Local Similarity 100.0%; Pred. No. 8.9e-172;
Matches 1090; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MAWKTLPIVILLLLSVFVIOQVSSQDLSSCAGRCGEGYSDATCNCYNCOHYMECCPDF 60
DB 1 MAWKTLPIVILLLLSVFVIOQVSSQDLSSCAGRCGEGYSDATCNCYNCOHYMECCPDF 60

QY 61 KRVCTAELSCGRCFESFERGECDCDAQCKYDKCCPDYEFCAEHSVSNSQSSSS 120
DB 61 KRVCTAELSCGRCFESFERGECDCDAQCKYDKCCPDYEFCAEHSVSNSQSSSS 120

QY 121 SSSSSSTTWK1KSSXNSANRELQKLVKXNKORTKKXTPPKPPVVDAGSLDNGD 180
DB 121 SSSSSSTTWK1KSSXNSANRELQKLVKXNKORTKKXTPPKPPVVDAGSLDNGD 180

QY 181 FKVTTPDSTTQHNKYSTSPKITTAKPINRPSLPNSDTSKETSITVNTKETVETKTT 240
DB 181 FKVTTPDSTTQHNKYSTSPKITTAKPINRPSLPNSDTSKETSITVNTKETVETKTT 240

QY 241 TTNKQSTGCKEKTTSAGTQSIKTSAXDLAPTSKVLAKPTPKAETTTKGPALTTPKEP 300
DB 241 TTNKQSTGCKEKTTSAGTQSIKTSAXDLAPTSKVLAKPTPKAETTTKGPALTTPKEP 300

QY 301 TPTPKPEASTTPKEPTPTTIKSAPTTPKEPAPTTTKSAPTTPKEPAPTTTPKEPAPT 360
DB 301 TPTPKPEASTTPKEPTPTTIKSAPTTPKEPAPTTTKSAPTTPKEPAPTTTPKEPAPT 360

QY 361 EPAPTTTKBPAPTTKSAPTTPKEPAPTTPKPAPTTTPKEPAPTTPKSPTPTTPKEPAPT 420
DB 361 EPAPTTTKBPAPTTKSAPTTPKEPAPTTPKPAPTTTPKEPAPTTPKSPTPTTPKEPAPT 420

QY 421 TKEPAPTTKEPAPTPAKPAPTPKEPAPTTPKPAPTTTPKEPAPTTTPKEPAPTTTSA 480
DB 421 TKEPAPTTKEPAPTPAKPAPTPKEPAPTTPKPAPTTTPKEPAPTTTPKEPAPTTTSA 480

QY 481 PTTTKEPAPTTKSAPTTPKEPAPTTTPKEPAPTTPKPAPTTTPKEPAPTTTPK 540
DB 481 PTTTKEPAPTTKSAPTTPKEPAPTTTPKEPAPTTPKPAPTTTPKEPAPTTTPK 540

QY 541 EPAPTTTKBPAPTPAKPAPTPKEPAPTTPKPAPTTTPKEPAPTTTPKEPAPTTPEELAP 600
DB 541 EPAPTTTKBPAPTPAKPAPTPKEPAPTTPKPAPTTTPKEPAPTTTPKEPAPTTPEELAP 600

QY 601 TTPPEPTPTTPBEAPPTPKAAPNTPKPAPTTTPKEPAPTTTPKEPAPTTTPKETAPTTPK 660
DB 601 TTPPEPTPTTPBEAPPTPKAAPNTPKPAPTTTPKEPAPTTTPKEPAPTTTPKETAPTTPK 660

QY 661 GTAPTTLKEPAPTTPKKAPKELAPTTTKEPTSTSDKAPAPTTTPKGAPTTPKBPAPTP 720
DB 661 GTAPTTLKEPAPTTPKKAPKELAPTTTKEPTSTSDKAPAPTTTPKGAPTTPKBPAPTP 720

QY 721 KEAPAPTTKGAPTTLKEPAPTTPKKAPKELAPTTTKEPTSTSDKAPAPTTTPKETAPTTP 780
DB 721 KEAPAPTTKGAPTTLKEPAPTTPKKAPKELAPTTTKEPTSTSDKAPAPTTTPKETAPTTP 780

QY 781 PKEPAPTTPKKAPAPTPPETPPPTSEVSTPTTKETPTTIHKSPDSESTPELSAETTPKALE 840
DB 781 PKEPAPTTPKKAPAPTPPETPPPTSEVSTPTTKETPTTIHKSPDSESTPELSAETTPKALE 840

QY 841 NSPKBPVPTTKTPAATKEMTTAKDKTTERDLATTPETTTAAPKMTKETATTEKTE 900
DB 841 NSPKBPVPTTKTPAATKEMTTAKDKTTERDLATTPETTTAAPKMTKETATTEKTE 900

QY 901 SKITATTTQVSTTTQDTPPKITLKTTLAPKVTITTKIITTEIMNKPEETAKPKDR 960
DB 901 SKITATTTQVSTTTQDTPPKITLKTTLAPKVTITTKIITTEIMNKPEETAKPKDR 960

QY 961 ATNSKATTPKQKPTKPKKPTSTKPKTMPVRKPTTPRKMSTMPBLNPTSIAE 1020

961 ATNSKATTPKQKPTKPKKPTSTKPKTMPVRKPTTPRKMSTMPBLNPTSIAE 1020

1021 AMLOTTTRPNQTPNSKLVEVNPKSEADAGAGETPHMLLRPHVFMPEVTPDMDYLPVFN 1080
1021 AMLOTTTRPNQTPNSKLVEVNPKSEADAGAGETPHMLLRPHVFMPEVTPDMDYLPVFN 1080

1081 QGIINPMLS 1090
1081 QGIINPMLS 1090

RESULT 2
US-07-757-022B-104
; Sequence 104, Application US/07757022B
; Patent No. 6433142
; GENERAL INFORMATION:
; APPLICANT: Gesner, Thomas G.
; APPLICANT: Clark, Stephen C.
; APPLICANT: Turner, Katherine
; APPLICANT: Hewick, Rodney M.
; TITLE OF INVENTION: Megakaryocyte Stimulating Factors
; NUMBER OF SEQUENCES: 143
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genetics Institute, Inc.
; STREET: 87 CambridgePark Drive
; CITY: Cambridge
; STATE: Massachusetts
; COUNTRY: U.S.A.
; ZIP: 02140
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/757,022B
; FILING DATE: 19910910
; CLASSIFICATION: 530
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/643,502
; FILING DATE: 18-JAN-1991
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/546,114
; FILING DATE: 29-JUN-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/457,196
; FILING DATE: 29-DEC-1989
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/390,901
; FILING DATE: 08-AUG-1989
; ATTORNEY/AGENT INFORMATION:
; NAME: Cserr, Luann
; REGISTRATION NUMBER: 31,822
; REFERENCE/DOCKET NUMBER: GI 5190
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617)876-1170
; TELEFAX: (617)876-5851
; INFORMATION FOR SEQ ID NO: 104:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1140 amino acids
; TYPE: AMINO ACID
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-07-757-022B-104

Query Match 99.7%; Score 5811; DB 4; Length 1140;
Best Local Similarity 95.6%; Pred. No. 2.1e-171;
Matches 1090; Conservative 0; Mismatches 0; Indels 50; Gaps 1;

QY 1 MAWKTLPIVILLLLSVFVIOQVSSQDLSSCAGRCGEGYSDATCNCYNCOHYMECCPDF 60
DB 1 MAWKTLPIVILLLLSVFVIOQVSSQDLSSCAGRCGEGYSDATCNCYNCOHYMECCPDF 60

QY 61 KRVCTAELSCGRCFESFERGECDCDAQCKKYDKCCPDYESFCA----- 105
Db 61 KRVCTAELSCGRCFESFERGECDCDAQCKKYDKCCPDYESFCAEVHNFTSPPPSKKAP 120
QY 106 -----FEHSVSENQSSSSSSSSSSSSSTI 130
Db 121 PPSGASQTIKSTTKSPKPNKKYKVKVIESBEITEHSVSENQSSSSSSSSSSSTI 180
QY 131 KIKSSKNSAANRELQKLVKONKNRKKKTPKPPVVVDEAGSLDNGDFKVTTPDTST 190
Db 181 KIKSSKNSAANRELQKLVKONKNRKKKTPKPPVVVDEAGSLDNGDFKVTTPDTST 240
QY 191 TOHNKVSTSPKLTAKPINPRESLPNSDTSKTSLSLVNKTETVETKTTINKOTSTDG 250
Db 241 TOHNKVSTSPKLTAKPINPRESLPNSDTSKTSLSLVNKTETVETKTTINKOTSTDG 300
QY 251 KEKTTSAKETQSIKTSADKLAPTSGVLAKPTPKAETTTKGPAITTPKEPTTTPKEPAS 310
Db 301 KEKTTSAKETQSIKTSADKLAPTSGVLAKPTPKAETTTKGPAITTPKEPTTTPKEPAS 360
QY 311 TTPKEPTTTPKSAPTTPKEPAITTPKSAPTTPKSAPTTPKSAPTTPKSAPTTPK 370
Db 361 TTPKEPTTTPKSAPTTPKEPAITTPKSAPTTPKSAPTTPKSAPTTPKSAPTTPK 420
QY 371 APTTTKSAPTTPKEPAITTPKSAPTTPKSAPTTPKSAPTTPKSAPTTPKSAPTTPK 430
Db 421 APTTTKSAPTTPKEPAITTPKSAPTTPKSAPTTPKSAPTTPKSAPTTPKSAPTTPK 480
QY 431 EPAPTAPKPAITTPKSAPTTPKSAPTTPKSAPTTPKSAPTTPKSAPTTPKSAPTTPK 490
Db 481 EPAPTAPKPAITTPKSAPTTPKSAPTTPKSAPTTPKSAPTTPKSAPTTPKSAPTTPK 540
QY 491 TTKSAPTTPKSPPTTPKEPAITTPKSAPTTPKSAPTTPKSAPTTPKSAPTTPKSAPTTPK 550
Db 541 TTKSAPTTPKSPPTTPKEPAITTPKSAPTTPKSAPTTPKSAPTTPKSAPTTPKSAPTTPK 600
QY 551 APTAPKEPAITTPKSAPTTPKSAPTTPKSAPTTPKSAPTTPKSAPTTPKSAPTTPK 610
Db 601 APTAPKEPAITTPKSAPTTPKSAPTTPKSAPTTPKSAPTTPKSAPTTPKSAPTTPK 660
QY 611 PEPAITTPKAAAPTTPKEPAITTPKSAPTTPKSAPTTPKSAPTTPKSAPTTPKSAPTTPK 670
Db 661 PEPAITTPKAAAPTTPKEPAITTPKSAPTTPKSAPTTPKSAPTTPKSAPTTPKSAPTTPK 720
QY 671 APTTPKPAKELAPITTPKSAPTTPKSAPTTPKSAPTTPKSAPTTPKSAPTTPKSAPTTPK 730
Db 721 APTTPKPAKELAPITTPKSAPTTPKSAPTTPKSAPTTPKSAPTTPKSAPTTPKSAPTTPK 780
QY 731 TAPTTLKSAPTTPKSAPTTPKSAPTTPKSAPTTPKSAPTTPKSAPTTPKSAPTTPKSAPTTPK 790
Db 781 TAPTTLKSAPTTPKSAPTTPKSAPTTPKSAPTTPKSAPTTPKSAPTTPKSAPTTPKSAPTTPK 840
QY 791 KEAPTTPPEPTTSEVSTPTTPKEPTTTPKSAPTTPKSAPTTPKSAPTTPKSAPTTPKSAPTTPK 850
Db 841 KEAPTTPPEPTTSEVSTPTTPKEPTTTPKSAPTTPKSAPTTPKSAPTTPKSAPTTPKSAPTTPK 900
QY 851 TKTPAATKPEMTTAKKITERDLRTTPETTTAAPTMTKETATTEKTTESKITATTITQV 910
Db 901 TKTPAATKPEMTTAKKITERDLRTTPETTTAAPTMTKETATTEKTTESKITATTITQV 960
QY 911 TSTTTQDTPPKITTLKTTLAPKVTITTKITTEIMNKPEEETAKPKDRATNSKATTPK 970
Db 961 TSTTTQDTPPKITTLKTTLAPKVTITTKITTEIMNKPEEETAKPKDRATNSKATTPK 1020
QY 971 FOKPTKAPKPTSTKPKTMTVRKPKTTPTPKMTSTMPNLNPTSRFAEAMLOTTTRPN 1030
Db 1021 FOKPTKAPKPTSTKPKTMTVRKPKTTPTPKMTSTMPNLNPTSRFAEAMLOTTTRPN 1080
QY 1031 QTPNSKLVEVPKPSDAGAGETPHMLLRPHVFNPEVTPDMDYLPVPNOGIIINPMLS 1090
Db 1081 QTPNSKLVEVPKPSDAGAGETPHMLLRPHVFNPEVTPDMDYLPVPNOGIIINPMLS 1140

RESULT 3

US-07-757-022B-2
; Sequence 2, Application US/07757022B
; Patent No. 6433142
; GENERAL INFORMATION:
; APPLICANT: Gesner, Thomas G.
; APPLICANT: Clark, Stephen C.
; APPLICANT: Turner, Katherine
; APPLICANT: Hewick, Rodney M.
; TITLE OF INVENTION: Megakaryocyte Stimulating Factors
; NUMBER OF SEQUENCES: 143
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genetics Institute, Inc.
; STREET: 87 Cambridgepark Drive
; CITY: Cambridge
; STATE: Massachusetts
; COUNTRY: U.S.A.
; ZIP: 02140
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/757,022B
; FILING DATE: 19910910
; CLASSIFICATION: 530
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/643,502
; FILING DATE: 18-JAN-1991
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/546,114
; FILING DATE: 29-JUN-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/457,196
; FILING DATE: 29-DEC-1989
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/390,901
; FILING DATE: 08-AUG-1989
; ATTORNEY/AGENT INFORMATION:
; NAME: Cseri, Luann
; REGISTRATION NUMBER: 31,822
; REFERENCE/DOCKET NUMBER: GI 5190
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617)876-1170
; TELEFAX: (617)876-5851
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1404 amino acids
; TYPE: AMINO ACID
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-07-757-022B-2

Query Match 99.7%; Score 5811; DB 4; Length 1404;
Best Local Similarity 95.6%; Pred. No. 2.7e-171;
Matches 1090; Conservative 0; Mismatches 0; Indels 50; Gaps 1;

QY 1 MAWKTPYILLLLSVFVIQVSSQDLSSCAGRCGEGYSDATCNCNDYNCQHYMECCPDF 60
Db 1 MAWKTPYILLLLSVFVIQVSSQDLSSCAGRCGEGYSDATCNCNDYNCQHYMECCPDF 60
QY 61 KRVCTAELSCGRCFESFERGECDCDAQCKKYDKCCPDYESFCA----- 105
Db 61 KRVCTAELSCGRCFESFERGECDCDAQCKKYDKCCPDYESFCAEVHNFTSPPPSKKAP 120
QY 106 -----FEHSVSENQSSSSSSSSSSSSSTI 130
Db 121 PPSGASQTIKSTTKSPKPNKKYKVKVIESBEITEHSVSENQSSSSSSSSSSSTI 180
QY 131 KIKSSKNSAANRELQKLVKONKNRKKKTPKPPVVVDEAGSLDNGDFKVTTPDTST 190

QY 1031 QTPNSKLVNPKSBDAGAGETPHMLLRPHVFMPEVTPDMVYLPVFNQGIINPMLS 1090
DB 1081 QTPNSKLVNPKSBDAGAGETPHMLLRPHVFMPEVTPDMVYLPVFNQGIINPMLS 1140

RESULT 6
US-10-164-595-78
; Sequence 78, Application US/10164595
; Patent No. 6657054
; GENERAL INFORMATION:
; APPLICANT: Origene Technologies, Inc
; TITLE OF INVENTION: Regulated Angiogenesis Genes and Polypeptides
; FILE REFERENCE: IU 103 R1
; CURRENT APPLICATION NUMBER: US/10/164,595
; CURRENT FILING DATE: 2002-06-10
; NUMBER OF SEQ ID NOS: 80
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 78
; LENGTH: 1404
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-164-595-78

Query Match 99.3%; Score 5788; DB 4; Length 1404;
Best Local Similarity 95.4%; Pred. No. 1.4e-170;
Matches 1087; Conservative 0; Mismatches 3; Indels 50; Gaps 1;

QY 1 MAWKTLPIYLLLSLVFVIQVSSODLSSCAGGEGYSRDATCNCYDNCQHYMECCPDF 60
DB 1 MAWKTLPIYLLLSLVFVIQVSSODLSSCAGGEGYSRDATCNCYDNCQHYMECCPDF 60

QY 61 KRVCTAELSCRCRCPESFERGECDCDAQCKYDKCCPDYSEFCA----- 105
DB 61 KRVCTAELSCRCRCPESFERGECDCDAQCKYDKCCPDYSEFCAEVNHNPTSPSSKAP 120

QY 106 -----EEHSVSENOESSSSSSSSSSSSSSSTIW 130
DB 121 PPSGASQTIKSTTKRSPXPNKKTKKVIIESEEITEHSVSENOESSSSSSSSSSSTIR 180

QY 131 KIKSSNSAANRELQKLVKNDKKNRTKPKTPPPVWDEAGSLGNDGPKVTPDTST 190
DB 181 KIKSSNSAANRELQKLVKNDKKNRTKPKTPPPVWDEAGSLGNDGPKVTPDTST 240

QY 191 TQHNKVSIPKTTTAKPNRPSLPNSDTSKETS LVNKEITVETKETTINKQTSIDG 250
DB 241 TQHNKVSIPKTTTAKPNRPSLPNSDTSKETS LVNKEITVETKETTINKQTSIDG 300

QY 251 KEKTSKAKETOSIEKTSADLAPTSKVLAKPTPKAEITTKGPAULTTPKEPTTPKPEPAS 310
DB 301 KEKTSKAKETOSIEKTSADLAPTSKVLAKPTPKAEITTKGPAULTTPKEPTTPKPEPAS 360

QY 311 TTPKEPTPTTKSAPTTKPEAPTTTKSAPTTKPEAPTTTKPEAPTTTKPEAPTTTKPE 370
DB 361 TTPKEPTPTTKSAPTTKPEAPTTTKSAPTTKPEAPTTTKPEAPTTTKPEAPTTTKPE 420

QY 371 APTTTKSAPTTKPEAPTTPKKAPTTTPKPEAPTTTPKPEPTTPKPEAPTTKPEAPTTPK 430
DB 421 APTTTKSAPTTKPEAPTTPKKAPTTTPKPEAPTTTPKPEPTTPKPEAPTTKPEAPTTPK 480

QY 431 EPAPTAPKAPTTKPEAPTTKPEAPTTTKESPTTPKPEAPTTTKSAPTTTKPEAPTT 490
DB 481 EPAPTAPKAPTTKPEAPTTKPEAPTTTKESPTTPKPEAPTTTKSAPTTTKPEAPTT 540

QY 491 TTKSAPTTKPESPPTTKPEAPTTPKKAPTTTPKPEAPTTTPKPEAPTTTKPEAPTTTKKP 550
DB 541 TTKSAPTTKPESPPTTKPEAPTTPKKAPTTTPKPEAPTTTPKPEAPTTTKPEAPTTTKKP 600

QY 551 APTAPKAPTTKPEAPTTKPEAPTTTPKPEAPTTTPKPEAPTTTPKPEAPTTTPKPEPTT 610
DB 601 APTTPKPEAPTTKPEAPTTTPKPEAPTTTPKPEAPTTTPKPEAPTTTPKPEAPTTTP 660

QY 611 PEEPAATPKAAAPNTPKEAPTTPKPEAPTTPKPEAPTTPKETAPTTPKGTAPTTLKEP 670

RESULT 7

US-07-757-022B-40
; Sequence 40, Application US/07757022B
; Patent No. 6433142
; GENERAL INFORMATION:
; APPLICANT: Gesner, Thomas G.
; APPLICANT: Clark, Stephen C.
; APPLICANT: Turner, Katherine
; APPLICANT: Hewick, Rodney M.
; TITLE OF INVENTION: Megakaryocyte Stimulating Factors
; NUMBER OF SEQUENCES: 143
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genetics Institute, Inc.
; STREET: 87 Cambridgepark Drive
; CITY: Cambridge
; STATE: Massachusetts
; COUNTRY: U.S.A.
; ZIP: 02140
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/757,022B
; FILING DATE: 19910910
; CLASSIFICATION: 530
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/643,502
; FILING DATE: 18-JAN-1991
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/546,114
; FILING DATE: 29-JUN-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/457,196
; FILING DATE: 29-DEC-1989
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/390,901
; FILING DATE: 08-AUG-1989
; ATTORNEY/AGENT INFORMATION:
; NAME: Cserr, Luann
; REGISTRATION NUMBER: 31,822


```
; REFERENCE/DOCKET NUMBER: GI 5190
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617)876-1170
; TELEFAX: (617)876-5851
; INFORMATION FOR SEQ ID NO: 40:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1361 amino acids
; TYPE: AMINO ACID
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-07-757-022B-40

Query Match          96.9%; Score 5643.3; DB 4; Length 1361;
Best Local Similarity 96.3%; Pred No. 3.8e-166;
Matches 1056; Conservative 8; Mismatches 26; Indels 7; Gaps 1;

QY 1 MAWKTLPIYLLLLSVFVIQVSSQDLSSCAGRCGEGYSDATNCNDCYNCHYMECCPDF 60
DB 1 MAWKTLPIYLLLLSVFVIQVSSQDLSSCAGRCGEGYSDATNCNDCYNCHYMECCPDF 60

QY 61 KVCVTAELSCRCRCEFERGECDCDAQCKYDKCCPDYSEFCAEHSESVSESSSS 120
DB 61 KVCVTAELSCRCRCEFERGECDCDAQCKYDKCCPDYSEFCAEHSESVSESSSS 120

QY 121 SSSSSSSITWIKSSKNGAANRELQKL-----KVKDNKNRRTKKKPTPKPPVVDAG 173
DB 121 SSSSSSSITWIKSSKNGAANRELQKL-----KVKDNKNRRTKKKPTPKPPVVDAG 173

QY 121 PPSGASQIKSTKESPPNNKKTKVIESBEITEVDKNKNRTKKKPTPKPPVVDAG 180
DB 121 PPSGASQIKSTKESPPNNKKTKVIESBEITEVDKNKNRTKKKPTPKPPVVDAG 180

QY 174 SGLDNGDKFVTPDSTSTQHNKVSTSPKITTAKPINRPSLPNPSDTSKETSIVNKEET 233
DB 174 SGLDNGDKFVTPDSTSTQHNKVSTSPKITTAKPINRPSLPNPSDTSKETSIVNKEET 233

QY 181 SGLDNGDKFVTPDSTSTQHNKVSTSPKITTAKPINRPSLPNPSDTSKETSIVNKEET 240
DB 181 SGLDNGDKFVTPDSTSTQHNKVSTSPKITTAKPINRPSLPNPSDTSKETSIVNKEET 240

QY 234 VETKETTNTNKTSTDGKEKTSKETSIAKTSKADLAPTSKVLAKPTPKAEATTIKGPA 293
DB 234 VETKETTNTNKTSTDGKEKTSKETSIAKTSKADLAPTSKVLAKPTPKAEATTIKGPA 293

QY 241 VETKETTNTNKTSTDGKEKTSKETSIAKTSKADLAPTSKVLAKPTPKAEATTIKGPA 300
DB 241 VETKETTNTNKTSTDGKEKTSKETSIAKTSKADLAPTSKVLAKPTPKAEATTIKGPA 300

QY 294 LPTTKEPTTTPKEPASTTPKEPTTTIKSAPTTKEPAPTTTKEPAPTTTKEPAPTTTKE 353
DB 294 LPTTKEPTTTPKEPASTTPKEPTTTIKSAPTTKEPAPTTTKEPAPTTTKEPAPTTTKE 353

QY 301 LPTTKEPTTTPKEPASTTPKEPTTTIKSAPTTKEPAPTTTKEPAPTTTKEPAPTTTKE 360
DB 301 LPTTKEPTTTPKEPASTTPKEPTTTIKSAPTTKEPAPTTTKEPAPTTTKEPAPTTTKE 360

QY 354 PAPTTPKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKE 413
DB 354 PAPTTPKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKE 413

QY 361 PAPTTPKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKE 420
DB 361 PAPTTPKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKE 420

QY 414 PKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKE 473
DB 414 PKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKE 473

QY 421 PKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKE 480
DB 421 PKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKE 480

QY 474 PTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKE 533
DB 474 PTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKE 533

QY 481 PTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKE 540
DB 481 PTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKE 540

QY 534 PAPTTPKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKE 593
DB 534 PAPTTPKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKE 593

QY 541 PAPTTPKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKE 600
DB 541 PAPTTPKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKE 600

QY 594 TPEELAPTTPEPTTTPKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKE 653
DB 594 TPEELAPTTPEPTTTPKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKE 653

QY 601 TPEELAPTTPEPTTTPKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKE 660
DB 601 TPEELAPTTPEPTTTPKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKE 660

QY 654 TAPTTPKGAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKE 713
DB 654 TAPTTPKGAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKE 713

QY 661 TAPTTPKGAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKE 720
DB 661 TAPTTPKGAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKE 720

QY 714 EPAPTTPKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKE 773
DB 714 EPAPTTPKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKE 773

QY 721 EPAPTTPKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKE 780
DB 721 EPAPTTPKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKE 780

QY 774 KETAPTTPKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKE 833
DB 774 KETAPTTPKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKE 833

QY 781 KETAPTTPKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKE 840
DB 781 KETAPTTPKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKE 840

QY 834 PTPKALENSPKPGVPTTKTPAAKPEMTTTAKDKTTERDLRTPETTTAAPTAKMTKETAT 893
DB 834 PTPKALENSPKPGVPTTKTPAAKPEMTTTAKDKTTERDLRTPETTTAAPTAKMTKETAT 893
```

RESULT 8

```
US-07-757-022B-42
; Sequence 42, Application US/07757022B
; Patent No. 6433142
; GENERAL INFORMATION:
; APPLICANT: Gesner, Thomas G.
; APPLICANT: Clark, Stephen C.
; APPLICANT: Turner, Katherine
; APPLICANT: Hewick, Rodney M.
; TITLE OF INVENTION: Megakaryocyte Stimulating Factors
; NUMBER OF SEQUENCES: 143
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genetics Institute, Inc.
; STREET: 87 CambridgePark Drive
; CITY: Cambridge
; STATE: Massachusetts
; COUNTRY: U.S.A.
; ZIP: 02140
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07757,022B
; FILING DATE: 19910910
; CLASSIFICATION: 530
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/643,502
; FILING DATE: 18-JAN-1991
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/546,114
; FILING DATE: 29-JUN-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/457,196
; FILING DATE: 29-DEC-1989
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/390,901
; FILING DATE: 08-AUG-1989
; ATTORNEY/AGENT INFORMATION:
; NAME: Cseri, Luann
; REGISTRATION NUMBER: 31,822
; REFERENCE/DOCKET NUMBER: GI 5190
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617)876-1170
; TELEFAX: (617)876-5851
; INFORMATION FOR SEQ ID NO: 42:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1361 amino acids
; TYPE: AMINO ACID
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-07-757-022B-42
```

Query Match 96.3%; Score 5608.7; DB 4; Length 1311;
Best Local Similarity 96.1%; Pred. No. 4.2e-165;
Matches 1047; Conservative 0; Mismatches 0; Indels 43; Gaps 1;

QY 1 MAWKTLPIYLLLSLVFVIQQVSSQDLSSCAGRGEGYSRDATCNCYNCQHYMECCPDF 60
DB 1 MAWKTLPIYLLLSLVFVIQQVSSQDLSSCAGRGEGYSRDATCNCYNCQHYMECCPDF 60

QY 61 KEVCTAELSCGRCPESPERGECDCDAQCKKYDKCCPDYSEFCAEBHSVSENGESSSS 120
DB 61 KEVCTAELSCGRCPESPERGECDCDAQCKKYDKCCPDYSEFCAEBHSVSENGESSSS 106

QY 121 SSSSSSTIWKI KSSKNSAANRELQKLVKNDKNRTKKXPTPKPPVVDAGSGLDNGD 180
DB 107 -----VKDNKNRTKKXPTPKPPVVDAGSGLDNGD 137

QY 181 FKVITPDSTTQHNKYSTSPKITTAKPINPRPSLPNSDTSKETSLSLWNETTVEKETT 240
DB 138 FKVITPDSTTQHNKYSTSPKITTAKPINPRPSLPNSDTSKETSLSLWNETTVEKETT 197

QY 241 TTNKOTSDGKEKITSAKETOSIEKTSADLAPTSKVLAKPTPKAETTTKGPALTTKPEP 300
DB 198 TTNKOTSDGKEKITSAKETOSIEKTSADLAPTSKVLAKPTPKAETTTKGPALTTKPEP 257

QY 301 TPTPKPEASTPKPEPTPTTKSAPTTKPEAPTTTKSAPTTKPEAPTTTKPEAPTTTK 360
DB 258 TPTPKPEASTPKPEPTPTTKSAPTTKPEAPTTTKSAPTTKPEAPTTTKPEAPTTTK 317

QY 361 EPAPTTKPEAPTTTKSAPTTKPEAPTTKPKAPTTKPEAPTTKPEPTTKPEAPT 420
DB 318 EPAPTTKPEAPTTTKSAPTTKPEAPTTKPKAPTTKPEAPTTKPEPTTKPEAPT 377

QY 421 TKEPAPTTKPEAPTTKPKAPTTKPEAPTTKPEAPTTTKPEPTTKPEAPTTKKSA 480
DB 378 TKEPAPTTKPEAPTTKPKAPTTKPEAPTTKPEAPTTTKPEPTTKPEAPTTKKSA 437

QY 481 PTTTKPEAPTTKSAPTTKPEPTTKPEAPTTKPEAPTTKPKAPTTKPEAPTTTK 540
DB 438 PTTTKPEAPTTKSAPTTKPEPTTKPEAPTTKPEAPTTKPKAPTTKPEAPTTTK 497

QY 541 EPAPTTTKKAPTAPEAPTTKPEPTTKPEAPTTKPEAPTTTKPEAPTTTKPEAP 600
DB 498 EPAPTTTKKAPTAPEAPTTKPEPTTKPEAPTTKPEAPTTTKPEAPTTTKPEAP 557

QY 601 TTPPEPTTKPEAPTTKPKAANTPKPEAPTTKPEAPTTKPEAPTTKPEAPTTTK 660
DB 558 TTPPEPTTKPEAPTTKPKAANTPKPEAPTTKPEAPTTKPEAPTTTKPEAPTTTK 617

QY 661 GTAPTTKPEAPTTKPKAPKELAPTTTKPEPTTKPEAPTTKPEAPTTTKPEAP 720
DB 618 GTAPTTKPEAPTTKPKAPKELAPTTTKPEPTTKPEAPTTKPEAPTTTKPEAP 677

QY 721 KEAPTTKPEAPTTKPEAPTTKPKAPKELAPTTTKGPTSTTSKPEAPTTKPEAPT 780
DB 678 KEAPTTKPEAPTTKPEAPTTKPKAPKELAPTTTKGPTSTTSKPEAPTTKPEAPT 737

QY 781 PKEAPTTKPEAPTTKPEPTTKPEPTTKPEPTTKPEPTTKPEPTTKPEPTTKPE 840
DB 738 PKEAPTTKPEAPTTKPEPTTKPEPTTKPEPTTKPEPTTKPEPTTKPEPTTKPE 797

QY 841 NSPKPEGVTTKPAATKPEMTTAKDKTTERDLRTTPEPTTKPEPTTKPEPTTKPE 900
DB 798 NSPKPEGVTTKPAATKPEMTTAKDKTTERDLRTTPEPTTKPEPTTKPEPTTKPE 857

QY 901 SKITATTTQVSTTQDTPPKITTLKTTLAPKVTITTKITTTTEIMNKEPTAKPKOR 960
DB 858 SKITATTTQVSTTQDTPPKITTLKTTLAPKVTITTKITTTTEIMNKEPTAKPKOR 917

QY 961 ATNSKATTPKPKAPKPTSTKPKTMRVRKPKTTTTPRQVSTMPPELNPSTRIAE 1020
DB 918 ATNSKATTPKPKAPKPTSTKPKTMRVRKPKTTTTPRQVSTMPPELNPSTRIAE 977

QY 1021 AMLQTTTRPNQTPNSKLVFNPKSBDAGAGETPHMLLRPHVFPVETPDMDYLPRVEN 1080
DB 978 AMLQTTTRPNQTPNSKLVFNPKSBDAGAGETPHMLLRPHVFPVETPDMDYLPRVEN 1037

QY 1081 QGIINPMLS 1090
DB 1038 QGIINPMLS 1047

RESULT 9
US-07-757-022B-50
; Sequence 50, Application US/07757022B
; Patent No. 6433142
; GENERAL INFORMATION:
; APPLICANT: Gesner, Thomas G.
; APPLICANT: Clark, Stephen C.
; APPLICANT: Turner, Katherine
; APPLICANT: Hewick, Rodney M.
; TITLE OF INVENTION: Megakaryocyte Stimulating Factors
; NUMBER OF SEQUENCES: 143
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genetics Institute, Inc.
; STREET: 87 CambridgePark Drive
; CITY: Cambridge
; STATE: Massachusetts
; COUNTRY: U.S.A.
; ZIP: 02140
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/757,022B
; FILING DATE: 19910910
; CLASSIFICATION: 530
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/643,502
; FILING DATE: 18-JAN-1991
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/546,114
; FILING DATE: 29-JUN-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/457,196
; FILING DATE: 29-DEC-1989
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/390,901
; FILING DATE: 08-AUG-1989
; ATTORNEY/AGENT INFORMATION:
; NAME: Cseri, Luann
; REGISTRATION NUMBER: 31,822
; REFERENCE/DOCKET NUMBER: GI 5190
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617)876-1170
; TELEFAX: (617)876-5851
; INFORMATION FOR SEQ ID NO: 50:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1314 amino acids
; TYPE: AMINO ACID
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-07-757-022B-50

Query Match 95.6%; Score 5571; DB 4; Length 1314;
Best Local Similarity 96.3%; Pred. No. 6e-164;
Matches 1050; Conservative 0; Mismatches 0; Indels 40; Gaps 1;

QY 1 MAWKTLPIYLLLSLVFVIQQVSSQDLSSCAGRGEGYSRDATCNCYNCQHYMECCPDF 60
DB 1 MAWKTLPIYLLLSLVFVIQQVSSQDLSSCAGRGEGYSRDATCNCYNCQHYMECCPDF 60

QY 61 KEVCTAELSCGRCPESPERGECDCDAQCKKYDKCCPDYSEFCAEBHSVSENGESSSS 120
DB 61 KEVCTAELSCGRCPESPERGECDCDAQCKKYDKCCPDYSEFCAEBHSVSENGESSSS 120

RESULT 10
US-07-757-

Db 200 TTNKQSTGDKKTTSAKTSQSIKTSKVLAKTSPKAEITTKGPAITTKP 259
Qy 301 TPTTKEPASTTKEPTTTTSAKTSKVLAKTSPKAEITTKGPAITTKP 360
Db 260 TPTTKEPASTTKEPTTTTSAKTSKVLAKTSPKAEITTKGPAITTKP 319
Qy 361 EPAPITTKKEPAPITTKSAPITTKKPAITTKKPAITTKKPAITTKP 420
Db 320 EPAPITTKKEPAPITTKSAPITTKKPAITTKKPAITTKKPAITTKP 379
Qy 421 TKEPAPITTKKEPAPITTKKPAITTKKPAITTKKPAITTKKPAITTKP 480
Db 380 TKEPAPITTKKEPAPITTKKPAITTKKPAITTKKPAITTKKPAITTKP 439
Qy 481 PTTTKEPAPITTKSAPITTKKPAITTKKPAITTKKPAITTKKPAITTKP 540
Db 440 PTTTKEPAPITTKSAPITTKKPAITTKKPAITTKKPAITTKKPAITTKP 499
Qy 541 EPAPITTKKEPAPITTKKPAITTKKPAITTKKPAITTKKPAITTKKPAITTKP 600
Db 500 EPAPITTKKEPAPITTKKPAITTKKPAITTKKPAITTKKPAITTKKPAITTKP 559
Qy 601 TPTTKEPAPITTKKPAITTKKPAITTKKPAITTKKPAITTKKPAITTKP 660
Db 560 TPTTKEPAPITTKKPAITTKKPAITTKKPAITTKKPAITTKKPAITTKP 619
Qy 661 GTAPITTKKEPAPITTKKPAITTKKPAITTKKPAITTKKPAITTKKPAITTKP 720
Db 620 GTAPITTKKEPAPITTKKPAITTKKPAITTKKPAITTKKPAITTKKPAITTKP 679
Qy 721 KEAPITTKKEPAPITTKKPAITTKKPAITTKKPAITTKKPAITTKKPAITTKP 780
Db 680 KEAPITTKKEPAPITTKKPAITTKKPAITTKKPAITTKKPAITTKKPAITTKP 739
Qy 781 PKEAPITTKKEPAPITTKKPAITTKKPAITTKKPAITTKKPAITTKKPAITTKP 840
Db 740 PKEAPITTKKEPAPITTKKPAITTKKPAITTKKPAITTKKPAITTKKPAITTKP 799
Qy 841 NSPKEPAPITTKKPAITTKKPAITTKKPAITTKKPAITTKKPAITTKKPAITTKP 900
Db 800 NSPKEPAPITTKKPAITTKKPAITTKKPAITTKKPAITTKKPAITTKKPAITTKP 859
Qy 901 SKITATITQVSTTQDTPPKITTKTTLAPKVTITTKITTTTTEIMNKPEETAKPKOR 960
Db 860 SKITATITQVSTTQDTPPKITTKTTLAPKVTITTKITTTTTEIMNKPEETAKPKOR 919
Qy 961 ATNSKATTPKPKKPAITTKKPAITTKKPAITTKKPAITTKKPAITTKKPAITTKP 1020
Db 920 ATNSKATTPKPKKPAITTKKPAITTKKPAITTKKPAITTKKPAITTKKPAITTKP 979
Qy 1021 AMLQITTRPNQTSKLVNPKSDAGGAGETPHMLLRPHVFMPEVTTPMDYLRVFN 1080
Db 980 AMLQITTRPNQTSKLVNPKSDAGGAGETPHMLLRPHVFMPEVTTPMDYLRVFN 1039
Qy 1081 QGIIINPMLS 1090
Db 1040 QGIIINPMLS 1049

RESULT 11

US-07-757-022B-142
; Sequence 142, Application US/07757022B
; Patent No. 6433142
; GENERAL INFORMATION:
; APPLICANT: Gesner, Thomas G.
; APPLICANT: Clark, Stephen C.
; APPLICANT: Turner, Katherine
; APPLICANT: Hewick, Rodney M.
; TITLE OF INVENTION: Megakaryocyte Stimulating Factors
; NUMBER OF SEQUENCES: 143
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genetics Institute, Inc.

STREET: 87 CambridgePark Drive
CITY: Cambridge
STATE: Massachusetts
COUNTRY: U.S.A.
ZIP: 02140
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/07/757,022B
FILING DATE: 19910910
CLASSIFICATION: 530
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/643,502
FILING DATE: 18-JAN-1991
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/546,114
FILING DATE: 29-JUN-1990
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/457,196
FILING DATE: 29-DEC-1989
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/390,901
FILING DATE: 08-AUG-1989
ATTORNEY/AGENT INFORMATION:
NAME: Cserr, Luann
REGISTRATION NUMBER: 31,822
REFERENCE/DOCKET NUMBER: GI 5190
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617)876-1170
TELEFAX: (617)876-5851
INFORMATION FOR SEQ ID NO: 142:
SEQUENCE CHARACTERISTICS:
LENGTH: 1313 amino acids
TYPE: AMINO ACID
TOPOLOGY: linear
MOLECULE TYPE: protein
US-07-757-022B-142

Query Match 95.5%; Score 5561.9; DB 4; Length 1313;
Best Local Similarity 96.2%; Pred. No. 1.2e-163;
Matches 1049; Conservative 0; Mismatches 0; Indels 41; Gaps 1;
Qy 1 MAWKTLPIYLLLLSVFVIQVSSQDLSSCAGRCGEGYSDATCNDYNCOHMECCPDF 60
Db 1 MAWKTLPIYLLLLSVFVIQVSSQ----- 25
Qy 61 KRVTAEISCKGRCPESFERGECDCDAQCKKYDKCCPDYSEFCAEHSVSESSSS 120
Db 26 -----ELSCGRCPESFERGECDCDAQCKKYDKCCPDYSEFCAEHSVSESSSS 79
Qy 121 SSSSSSTIWKIKSKNSAANRELQKLKVXDNKKNRTKKKPTPKPPVVDAGSLDNGD 180
Db 80 SSSSSSTIWKIKSKNSAANRELQKLKVXDNKKNRTKKKPTPKPPVVDAGSLDNGD 139
Qy 181 FKVTPDSTTCHNKVSTSPKITTAKPINRPSLPNPSDTSKETSIVNKETTVETKETT 240
Db 140 FKVTPDSTTCHNKVSTSPKITTAKPINRPSLPNPSDTSKETSIVNKETTVETKETT 199
Qy 241 TTNKQSTGDKKTTSAKTSQSIKTSKVLAKTSPKAEITTKGPAITTKP 300
Db 200 TTNKQSTGDKKTTSAKTSQSIKTSKVLAKTSPKAEITTKGPAITTKP 259
Qy 301 TPTTKEPASTTKEPTTTTSAKTSKVLAKTSPKAEITTKGPAITTKP 360
Db 260 TPTTKEPASTTKEPTTTTSAKTSKVLAKTSPKAEITTKGPAITTKP 319
Qy 361 EPAPITTKKEPAPITTKSAPITTKKPAITTKKPAITTKKPAITTKKPAITTKP 420
Db 320 EPAPITTKKEPAPITTKSAPITTKKPAITTKKPAITTKKPAITTKKPAITTKP 379

551 APTAPKEPAPTTKETAPTTKLTPTTPEKLAFTTPEKPAFTTPEELAPTTPEEPTPTT 610
560 APTAPKEPAPTTKETAPTTKLTPTTPEKLAFTTPEKPAFTTPEELAPTTPEEPTPTT 619
611 PEPAPPTPKAANPTPEKPAFTTPEKPAFTTPEKPAFTTPEKPAFTTPEKPAFTTPEK 670
620 PEPAPPTPKAANPTPEKPAFTTPEKPAFTTPEKPAFTTPEKPAFTTPEKPAFTTPEK 679
671 APTPEKPAKELAPTTKETPTSTSDKPAFTTPEKPAFTTPEKPAFTTPEKPAFTTPEK 730
680 APTPEKPAKELAPTTKETPTSTSDKPAFTTPEKPAFTTPEKPAFTTPEKPAFTTPEK 739
731 TAPTTLKEPAPTTKPAKPAKELAPTTKGTSTTSKPAFTTPEKPAFTTPEKPAFTTPEK 790
740 TAPTTLKEPAPTTKPAKPAKELAPTTKGTSTTSKPAFTTPEKPAFTTPEKPAFTTPEK 799
791 KPAFTTPEPPTTSVSTPTTKEPTTIHKSDESTPELSAFTTPEKPAFTTPEKPAFTTPEK 850
800 KPAFTTPEPPTTSVSTPTTKEPTTIHKSDESTPELSAFTTPEKPAFTTPEKPAFTTPEK 859
851 TKTPAATKPMITAKDKITERDLRTTPTTAAKPMITAKTETATTTETTESKITATTQV 910
860 TKTPAATKPMITAKDKITERDLRTTPTTAAKPMITAKTETATTTETTESKITATTQV 919
911 TSTTTQDTPFKITLKTITLAPKVTTKTITTTIIMKPBETAKPKDRATNSKATTPK 970
920 TSTTTQDTPFKITLKTITLAPKVTTKTITTTIIMKPBETAKPKDRATNSKATTPK 979
971 POKPTAPKPTSTKPKMTPVRKPTTPTPKMTSTWPNLPTSRIAEAMLOTTTPN 1030
980 POKPTAPKPTSTKPKMTPVRKPTTPTPKMTSTWPNLPTSRIAEAMLOTTTPN 1039
1031 QTPNSKLVEVPKSDAGAGETPHMLLRPHVFMPEVTPDMDYLPVFNQGIINPMLS 1090
1040 QTPNSKLVEVPKSDAGAGETPHMLLRPHVFMPEVTPDMDYLPVFNQGIINPMLS 1099

RESULT 13

US-07-757-022B-84
; Sequence 84, Application US/07/757022B
; Patent No. 6433142
; GENERAL INFORMATION:
; APPLICANT: Gesner, Thomas G.
; APPLICANT: Clark, Stephen C.
; APPLICANT: Turner, Katherine
; APPLICANT: Hewick, Rodney M.
; APPLICANT: Genetics Institute, Inc.
; STREET: 87 CambridgePark Drive
; CITY: Cambridge
; STATE: Massachusetts
; COUNTRY: U.S.A.
; ZIP: 02140
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/757,022B
; FILING DATE: 19910910
; CLASSIFICATION: 530
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/643,502
; FILING DATE: 18-JAN-1991
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/546,114
; FILING DATE: 29-JUN-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/457,196

; FILING DATE: 29-DEC-1989
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/390,901
; FILING DATE: 08-AUG-1989
; ATTORNEY/AGENT INFORMATION:
; NAME: Cseir, Luann
; REGISTRATION NUMBER: 31,822
; REFERENCE/DOCKET NUMBER: GI 5190
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617)876-1170
; TELEFAX: (617)876-5851
; INFORMATION FOR SEQ ID NO: 84:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1022 amino acids
; TYPE: AMINO ACID
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-07-757-022B-84

Query Match 93.8%; Score 5466.7; DB 4; Length 1022;
Best Local Similarity 95.8%; Pred. No. 7.1e-161;
Matches 1020; Conservative 0; Mismatches 2; Indels 43; Gaps 2;

QY 26 DLSSCAGRCGEGYSRDATCNCYNCOHYMECCDFKVCVCTAEISCKGRCFESERREC 85
DB 1 DLSSCAGRCGEGYSRDATCNCYNCOHYMECCDFKVC--ELSCGRCFESERREC 58
QY 86 CDAQCKKYDKCCPDYSEFCAEHSVSENQSSSSSSSSSSSSSSSIWIKSKNSAANRELQ 145
DB 59 CDAQCKKYDKCCPDYSEFCAE----- 79
QY 146 KKLKVKONKORPKKKTTPKPPVVDAGSGLDNGDPKVTTPDSTTOHKNKVSPTKITT 205
DB 80 --TAVKONKORPKKKTTPKPPVVDAGSGLDNGDPKVTTPDSTTQHNKVSPTKITT 137
QY 206 KPINRPSLPNSDTSKETSITVTKETTTNKQTSIDGKEKTTSAKETQSIK 265
DB 138 KPINRPSLPNSDTSKETSITVTKETTTNKQTSIDGKEKTTSAKETQSIK 197
QY 266 TSAKDLAPTSKVLAKEPTTKEPAFTTTPKEPTTTPKEPAFTTTPKEPTTTPKEPTTTP 325
DB 198 TSAKDLAPTSKVLAKEPTTKEPAFTTTPKEPTTTPKEPAFTTTPKEPTTTPKEPTTTP 257
QY 326 TTPKEPAFTTTPKEPAFTTTPKEPAFTTTPKEPAFTTTPKEPAFTTTPKEPAFTTTPKE 385
DB 258 TTPKEPAFTTTPKEPAFTTTPKEPAFTTTPKEPAFTTTPKEPAFTTTPKEPAFTTTPKE 317
QY 386 APTTPKPAFTTTPKEPAFTTTPKEPAFTTTPKEPAFTTTPKEPAFTTTPKEPAFTTTP 445
DB 318 APTTPKPAFTTTPKEPAFTTTPKEPAFTTTPKEPAFTTTPKEPAFTTTPKEPAFTTTP 377
QY 446 KEPAFTTTPKEPAFTTTPKEPAFTTTPKEPAFTTTPKEPAFTTTPKEPAFTTTPKEPSPT 505
DB 378 KEPAFTTTPKEPAFTTTPKEPAFTTTPKEPAFTTTPKEPAFTTTPKEPAFTTTPKEPSPT 437
QY 506 TTKEPAFTTTPKEPAFTTTPKEPAFTTTPKEPAFTTTPKEPAFTTTPKEPAFTTTPKE 565
DB 438 TTKEPAFTTTPKEPAFTTTPKEPAFTTTPKEPAFTTTPKEPAFTTTPKEPAFTTTPKE 497
QY 566 TAPTPKLTPTTPEKLAFTTPEKPAFTTPEELAPTTPEEPTPTTPEEPTPTTPEEPTPTT 625
DB 498 TAPTPKLTPTTPEKLAFTTPEKLAFTTPEELAPTTPEEPTPTTPEEPTPTTPEEPTPTT 557
QY 626 TPKEPAFTTTPKEPAFTTTPKEPAFTTTPKEPAFTTTPKEPAFTTTPKEPAFTTTPKE 685
DB 558 TPKEPAFTTTPKEPAFTTTPKEPAFTTTPKEPAFTTTPKEPAFTTTPKEPAFTTTPKE 617
QY 686 TTTKEPTSTTSKPAFTTTPKGTATTPKPAFTTTPKEPAFTTTPKGTATTPKPAFTTTPK 745
DB 618 TTTKEPTSTTSKPAFTTTPKGTATTPKPAFTTTPKEPAFTTTPKGTATTPKPAFTTTPK 677
QY 746 KPAKELAPTTTKGPTSTTSDKPAFTTTPKEPAFTTTPKEPAFTTTPKEPAFTTTPKEPTT 805

Db 920 TAKPKDRATNSKATTPKQKP:KAPKPTSTCKPKTMRVAKPTTTPRKTSTMPDLN 979
Qy 1014 PISRIAEAMLQTTTPNQTPNSKLVNPKSDAGAGETPHMLLRPHVFMPEVTPDMD 1073
Db 980 PISRIAEAMLQTTTPNQTPNSKLVNPKSDAGAGETPHMLLRPHVFMPEVTPDMD 1039
Qy 1074 YLPRVFNQGIINPMLS 1090
Db 1040 YLPRVFNQGIINPMLS 1056

RESULT 15

US-07-757-022B-60
; Sequence 60, Application US/07757022B
; Patent No. 6433142
; GENERAL INFORMATION:
; APPLICANT: Gesner, Thomas G.
; APPLICANT: Clark, Stephen C.
; APPLICANT: Turner, Katherine
; APPLICANT: Hewick, Rodney M.
; TITLE OF INVENTION: Megakaryocyte Stimulating Factors
; NUMBER OF SEQUENCES: 143
; CORRESPONDENCE ADDRESS: 143
; ADDRESSEE: Genetics Institute, Inc.
; STREET: 87 Cambridgepark Drive
; CITY: Cambridge
; STATE: Massachusetts
; COUNTRY: U.S.A.
; ZIP: 02140
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/757,022B
; FILING DATE: 19910910
; CLASSIFICATION: 530
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/643,502
; FILING DATE: 18-JAN-1991
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/546,114
; FILING DATE: 29-JUN-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/457,196
; FILING DATE: 29-DEC-1989
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/390,901
; FILING DATE: 08-AUG-1989
; ATTORNEY/AGENT INFORMATION:
; NAME: Cseerr, Luann
; REGISTRATION NUMBER: 31,822
; REFERENCE/DOCKET NUMBER: GI 5190
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617)876-1170
; TELEFAX: (617)876-5851
; INFORMATION FOR SEQ ID NO: 60:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1320 amino acids
; TYPE: AMINO ACID
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-07-757-022B-60

Query Match 92.3%; Score 5379.2; DB 4; Length 1320;
Best Local Similarity 92.5%; Pred. No. 4.9e-158;
Matches 1015; Conservative 8; Mismatches 26; Indels 48; Gaps 2;
Qy 1 MAWKLPYLLLLSVFVVIQVSSODLSSCAGRCGEGYSDATCNCNDYNCQHYMECCPDF 60
Db 1 MAWKLPYLLLLSVFVVIQVSSODLSSCAGRCGEGYSDATCNCNDYNCQHYMECCPDF 25

Qy 61 KRYCTAELSCKGRCPESFERGREGDCDAOCKKYDKCCPDYESFCABEHSVSSENQSSSS 120
Db 26 -----ELSCKGRCPESFERGREGDCDAOCKKYDKCCPDYESFCABEVNPTSPFSSKAP 79
Qy 121 SSSSSSTIWKIKSSKNXSAANRELQKL-----XVDNKKXRTKKKTPKPPVVDKAG 173
Db 80 PPSGASQTIKSTTKRSPKPPNKKTKKVIESBEITEVDKNDKXRTKKKTPKPPVVDKAG 139
Qy 174 SGLDNGDFKVTTPDSTTTOHNVKSTSPKITTAKPINRPSLPPNSDTSKETSITVKNKETT 233
Db 140 SGLDNGDFKVTTPDSTTTOHNVKSTSPKITTAKPINRPSLPPNSDTSKETSITVKNKETT 199
Qy 234 VETKETTNTNKQTSDDGKEKTTSAKETQSIETSAKDLPAPTSKVLAKPTPKAETTTKGA 293
Db 200 VETKETTNTNKQTSDDGKEKTTSAKETQSIETSAKDLPAPTSKVLAKPTPKAETTTKGA 259
Qy 294 LTTPKETPTTPKEPASTTPKEPTPTTIKSAPTTKEPAPTTTKSAPTTKEPAPTTTKE 353
Db 260 LTTPKETPTTPKEPASTTPKEPTPTTIKSAPTTKEPAPTTTKSAPTTKEPAPTTTKE 319
Qy 354 PAPTTPKEPAPTTTKEPAPTTTTSAPTTTPKEPAPTTTKKAPATTTPKEPAPTTTKEPTPTT 413
Db 320 PAPTTPKEPAPTTTKEPAPTTTTSAPTTTPKEPAPTTTKKAPATTTPKEPAPTTTKEPTPTT 379
Qy 414 PKEPAPTTKEPAPTTTPKEPAPTAAPKAPADTTTPKEPAPTTTPKEPAPTTTKEPAPTTTKE 473
Db 380 PKEPAPTTKEPAPTTTPKEPAPTAAPKAPADTTTPKEPAPTTTPKEPAPTTTKEPAPTTTKE 439
Qy 474 PTTTTSAPTTTKEPAPTTTTSAPTTTPKEPAPTTTTPKEPAPTTTTPKEPAPTTTTPKE 533
Db 440 PTTTTSAPTTTKEPAPTTTTSAPTTTPKEPAPTTTTPKEPAPTTTTPKEPAPTTTTPKE 499
Qy 534 PAPTTPKEPAPTTTTPKEPAPTAAPKAPADTTTPKEPAPTTTTPKEPAPTTTTPKEPAPTTTTPKE 593
Db 500 PAPTTPKEPAPTTTTPKEPAPTAAPKAPADTTTPKEPAPTTTTPKEPAPTTTTPKEPAPTTTTPKE 559
Qy 594 TPPELAPTTPEEPTPTTPPEPAPTTPKAAAPNTPKEPAPTTTPKEPAPTTTPKEPAPTTTKE 653
Db 560 TPPELAPTTPEEPTPTTPPEPAPTTPKAAAPNTPKEPAPTTTPKEPAPTTTPKEPAPTTTKE 619
Qy 654 TAPTTPKGATPTTLKEPAPTTTPKKAPKELAPTTTKEPTSTTSKAPATTTPKGATPTTPK 713
Db 620 TAPTTPKGATPTTLKEPAPTTTPKKAPKELAPTTTKEPTSTTSKAPATTTPKGATPTTPK 679
Qy 714 EPAPTTPKEPAPTTTPKGATPTTLKEPAPTTTPKKAPKELAPTTTKEPTSTTSKAPATTTPK 773
Db 680 EPAPTTPKEPAPTTTPKGATPTTLKEPAPTTTPKKAPKELAPTTTKEPTSTTSKAPATTTPK 739
Qy 774 KETAPTTTPKEPAPTTTPKEPAPTTTPPEPTTSEVSTPTTKEPTTIHKSPDSESTPELSAE 833
Db 740 KETAPTTTPKEPAPTTTPKEPAPTTTPPEPTTSEVSTPTTKEPTTIHKSPDSESTPELSAE 799
Qy 834 PTPKALENSPKBEGVPTTKTAAATKPEMTTAKDKTTERDLRTTTPETTTAAAPKMTKETAT 893
Db 800 PTPKALENSPKBEGVPTTKTAAATKPEMTTAKDKTTERDLRTTTPETTTAAAPKMTKETAT 859
Qy 894 TTEKTTESKITATTTTQVSTTTQDTPPKITTLKTTTLAPKVTITTKTITTTTETIMNKPEE 953
Db 860 TTEKTTESKITATTTTQVSTTTQDTPPKITTLKTTTLAPKVTITTKTITTTTETIMNKPEE 919
Qy 954 TAKPKDRATNSKATTPKQKPKAPKPTSTTKKPKTMPVRKPKTTPTRKMTSTMPDLN 1013
Db 920 TAKPKDRATNSKATTPKQKPKAPKPTSTTKKPKTMPVRKPKTTPTRKMTSTMPDLN 979
Qy 1014 PISRIAEAMLQTTTPNQTPNSKLVNPKSDAGAGETPHMLLRPHVFMPEVTPDMD 1073
Db 980 PISRIAEAMLQTTTPNQTPNSKLVNPKSDAGAGETPHMLLRPHVFMPEVTPDMD 1039
Qy 1074 YLPRVFNQGIINPMLS 1090
Db 1040 YLPRVFNQGIINPMLS 1056

Search completed: October 13, 2004, 11:58:50
JOB time : 23.9474 secs

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: October 13, 2004, 11:23:49 ; Search time 110.895 Seconds
(without alignments)
3171.696 Million cell updates/sec

Title: SEQ1-B

Perfect score: 5826

Sequence: 1 MAWTKLPYLLLSVFVIQ.....DMDYLPVPGIINPMLS 1090

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.1

Searched: 1356558 seqs, 322682953 residues

Total number of hits satisfying chosen parameters: 1356558

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications AA:*

1: /cgn2_6/ptodata/1/pubpaa/US07_PUBCOMB.pep.*
2: /cgn2_6/ptodata/1/pubpaa/PCT_NEW_PUB.pep.*
3: /cgn2_6/ptodata/1/pubpaa/US06_NEW_PUB.pep.*
4: /cgn2_6/ptodata/1/pubpaa/US06_PUBCOMB.pep.*
5: /cgn2_6/ptodata/1/pubpaa/US07_NEW_PUB.pep.*
6: /cgn2_6/ptodata/1/pubpaa/PCTUS_PUBCOMB.pep.*
7: /cgn2_6/ptodata/1/pubpaa/US08_NEW_PUB.pep.*
8: /cgn2_6/ptodata/1/pubpaa/US08_PUBCOMB.pep.*
9: /cgn2_6/ptodata/1/pubpaa/US09A_PUBCOMB.pep.*
10: /cgn2_6/ptodata/1/pubpaa/US09B_PUBCOMB.pep.*
11: /cgn2_6/ptodata/1/pubpaa/US09C_PUBCOMB.pep.*
12: /cgn2_6/ptodata/1/pubpaa/US09_NEW_PUB.pep.*
13: /cgn2_6/ptodata/1/pubpaa/US10A_PUBCOMB.pep.*
14: /cgn2_6/ptodata/1/pubpaa/US10B_PUBCOMB.pep.*
15: /cgn2_6/ptodata/1/pubpaa/US10C_PUBCOMB.pep.*
16: /cgn2_6/ptodata/1/pubpaa/US10D_PUBCOMB.pep.*
17: /cgn2_6/ptodata/1/pubpaa/US10_NEW_PUB.pep.*
18: /cgn2_6/ptodata/1/pubpaa/US60_NEW_PUB.pep.*
19: /cgn2_6/ptodata/1/pubpaa/US60_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | ID | Description |
|------------|--------|-------------|--------|----|-------------------|
| 1 | 5826 | 100.0 | 1354 | 13 | US-10-124-557-48 |
| 2 | 5811 | 93.7 | 1140 | 13 | US-10-124-557-104 |
| 3 | 5811 | 93.7 | 1404 | 9 | US-09-802-207-30 |
| 4 | 5811 | 93.7 | 1404 | 11 | US-09-897-188-1 |
| 5 | 5811 | 93.7 | 1404 | 13 | US-10-124-557-2 |
| 6 | 5811 | 93.7 | 1404 | 13 | US-10-124-557-62 |
| 7 | 5643.3 | 96.9 | 1361 | 13 | US-10-124-557-40 |
| 8 | 5608.7 | 96.3 | 1311 | 13 | US-10-124-557-42 |
| 9 | 5571 | 95.6 | 1314 | 13 | US-10-124-557-50 |
| 10 | 5561.9 | 95.5 | 1049 | 13 | US-10-124-557-58 |
| 11 | 5561.9 | 95.5 | 1313 | 13 | US-10-124-557-142 |
| 12 | 5546.9 | 95.2 | 1363 | 13 | US-10-124-557-52 |
| 13 | 5466.7 | 93.8 | 1022 | 13 | US-10-124-557-84 |
| 14 | 5373.2 | 92.3 | 1320 | 13 | US-10-124-557-46 |

15 5379.2 92.3 1320 13 US-10-124-557-60
16 5344.6 91.7 1038 13 US-10-124-557-74
17 5344.6 91.7 1270 13 US-10-124-557-44
18 5011 86.0 941 13 US-10-124-557-14
19 2446.8 42.0 732 9 US-09-802-207-27
20 1409.9 24.2 538 14 US-10-038-694-3
21 1394.5 23.9 5179 9 US-09-922-217-1068
22 1394.5 23.9 5179 9 US-09-833-263-1068
23 1394.5 23.9 5179 13 US-10-025-380-1068
24 1394.5 23.9 5179 16 US-10-734-564-111
25 1335.4 22.9 232 16 US-10-468-910-4
26 1010.7 17.3 188 14 US-10-038-694-2
27 960.8 16.5 1460 14 US-10-295-027-428
28 938.9 16.1 1325 9 US-09-864-761-35612
29 928.3 15.9 1367 9 US-09-801-368-108
30 914 15.7 220 13 US-10-124-557-96
31 904.3 15.5 207 13 US-10-124-557-116
32 904.3 15.5 207 13 US-10-124-557-136
33 901.5 15.3 3507 14 US-10-369-493-5784
34 869.7 14.9 157 13 US-10-124-557-102
35 869.7 14.9 157 13 US-10-124-557-114
36 862.5 14.8 6642 14 US-10-369-493-5013
37 859 14.7 2090 16 US-10-408-765A-2318
38 852 14.6 203 13 US-10-124-557-94
39 847 14.5 208 13 US-10-124-557-132
40 839.5 14.4 3256 16 US-10-408-765A-174
41 839.5 14.4 3256 16 US-10-701-490-9
42 838.2 14.4 463 13 US-10-124-557-54
43 836 14.3 19723 15 US-10-084-846A-5
44 832.7 14.3 697 15 US-10-425-114-41545
45 831.7 14.3 3256 9 US-09-919-172-99

ALIGNMENTS

RESULT 1

US-10-124-557-48
; Sequence 48, Application US/10124557
; Publication NO. US20020137894A1
; GENERAL INFORMATION:
APPLICANT: Turner, Katherine
Clark, Stephen C.
Jacobs, Kenneth M.
Hewick, Rodney M.
Gesser, Thomas M.
TITLE OF INVENTION: Megakaryocyte Stimulating Factors
NUMBER OF SEQUENCES: 143
CORRESPONDENCE ADDRESS:
ADDRESSEE: Genetics Institute, Inc.
STREET: 87 CambridgePark Drive
CITY: Cambridge
STATE: Massachusetts
COUNTRY: U.S.A.
ZIP: 02140
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION NUMBER: US/10/124,557
FILING DATE: 16-Apr-2002
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/643,502
FILING DATE: 18-JAN-1991
APPLICATION NUMBER: US 07/546,114
FILING DATE: 29-JUN-1990
APPLICATION NUMBER: US 07/457,196
FILING DATE: 29-DEC-1989
APPLICATION NUMBER: US 07/390,901
FILING DATE: 08-AUG-1989

```

ATTORNEY/AGENT INFORMATION:
NAME: Cserr, Luann
REGISTRATION NUMBER: 31,822
REFERENCE/DOCKET NUMBER: GI 5190
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617)876-1170
TELEFAX: (617)876-5851
INFORMATION FOR SEQ ID NO: 48:
SEQUENCE CHARACTERISTICS:
LENGTH: 1354 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
SEQUENCE DESCRIPTION: SEQ ID NO: 48:
US-10-124-557-48

Query Match      100.0%; Score 5826; DB 13; Length 1354;
Best Local Similarity 100.0%; Pred. No. 2.6e-143;
Matches 1090; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MAWKTLPIYLILLISLVFVIQVSSQDISSCAGRCGEGYSRDATCNCYNCHYMECCPDF 60
DB 1 MAWKTLPIYLILLISLVFVIQVSSQDISSCAGRCGEGYSRDATCNCYNCHYMECCPDF 60
QY 61 KRVTAEISCKGRGCFESFERGECDDAQCCKYDKCCPDYEFSCABEHSVSENQESSSS 120
DB 61 KRVTAEISCKGRGCFESFERGECDDAQCCKYDKCCPDYEFSCABEHSVSENQESSSS 120
QY 121 SSSSSSTIWKIKSSKNSAANRELQKLVKDNKKNKTKKPPKPPVVDVDEAGSLDNGD 180
DB 121 SSSSSSTIWKIKSSKNSAANRELQKLVKDNKKNKTKKPPKPPVVDVDEAGSLDNGD 180
QY 181 FKVTEDSTTHQNKVSTSPKLTAKPINRPSPSPNSDTSKETSLSLVNKETTVETKETT 240
DB 181 FKVTEDSTTHQNKVSTSPKLTAKPINRPSPSPNSDTSKETSLSLVNKETTVETKETT 240
QY 241 TTNKQSTDGKEKTTSAKETOSIKTSKADLAPTSKVLAKPTPKAETTTKGPALTTPKPEP 300
DB 241 TTNKQSTDGKEKTTSAKETOSIKTSKADLAPTSKVLAKPTPKAETTTKGPALTTPKPEP 300
QY 301 TPPTPKPEASTTPKEPTTTIKSAPTTPKAPPTTKSAPTTKPEAPTTKPEAPTTKPEAPTT 360
DB 301 TPPTPKPEASTTPKEPTTTIKSAPTTPKAPPTTKSAPTTKPEAPTTKPEAPTTKPEAPTT 360
QY 361 EPAPTTTKEAPATTTKSAPTTPKAPPTTKKAPPTTKPEAPTTKPEAPTTKPEAPTTKPEAPT 420
DB 361 EPAPTTTKEAPATTTKSAPTTPKAPPTTKKAPPTTKPEAPTTKPEAPTTKPEAPTTKPEAPT 420
QY 421 TKEPAPTTKPEAPATPKKAPPTTKPEAPTTKPEAPTTKPEAPTTKPEAPTTKPEAPTTKSA 480
DB 421 TKEPAPTTKPEAPATPKKAPPTTKPEAPTTKPEAPTTKPEAPTTKPEAPTTKPEAPTTKSA 480
QY 481 PTTKKEPAPTTTKGAPTTKPEAPTTTKEPAPTTKPEAPTTKPEAPTTKPEAPTTKPEAPTT 540
DB 481 PTTKKEPAPTTTKGAPTTKPEAPTTTKEPAPTTKPEAPTTKPEAPTTKPEAPTTKPEAPTT 540
QY 541 EPAPTTTKKAPATPKAPKAPPTTKETAPTTKLTPTTPEKLAAPTTPKPEAPTTPEELAP 600
DB 541 EPAPTTTKKAPATPKAPKAPPTTKETAPTTKLTPTTPEKLAAPTTPKPEAPTTPEELAP 600
QY 601 TTPPEPTTTPEEAPPTPKAAANTPKPEAPTTKPEAPTTKPEAPTTKPEAPTTKPEAPTTK 660
DB 601 TTPPEPTTTPEEAPPTPKAAANTPKPEAPTTKPEAPTTKPEAPTTKPEAPTTKPEAPTTK 660
QY 661 GTAPTTLKEPAPTTPKKAPKELAPTTTKEPTSTTSKAPPTPKGAPTTKPEAPTTTP 720
DB 661 GTAPTTLKEPAPTTPKKAPKELAPTTTKEPTSTTSKAPPTPKGAPTTKPEAPTTTP 720
QY 721 KEAPATTPKGTAPTTLKEPAPTTPKKAPKELAPTTTKGPTSTTSKAPPTTKETAPTT 780
DB 721 KEAPATTPKGTAPTTLKEPAPTTPKKAPKELAPTTTKGPTSTTSKAPPTTKETAPTT 780
QY 781 PKEAPATTPKKAPPTTPTPTPTTSEVSTPTTTKEPTTIHKSPDESPPELSAETPKALE 840

```

RESULT 2

US-10-124-557-104
 : Sequence 104, Application US/10124557
 : Publication No. US20020137894A1

GENERAL INFORMATION:

APPLICANT: Turner, Katherine
 Clark, Stephen C.
 Jacobs, Kenneth
 Hewick, Rodney M.
 Gesner, Thomas G.

TITLE OF INVENTION: Megakaryocyte Stimulating Factors

NUMBER OF SEQUENCES: 143

CORRESPONDENCE ADDRESS:

ADDRESSEE: Genetics Institute, Inc.
 STREET: 87 Cambridgepark Drive
 CITY: Cambridge
 STATE: Massachusetts
 COUNTRY: U.S.A.
 ZIP: 02140

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: Patent In Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/10/124,557
 FILING DATE: 16-Apr-2002
 CLASSIFICATION: <Unknown>

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 07/643,502
 FILING DATE: 18-JAN-1991
 APPLICATION NUMBER: US 07/546,114
 FILING DATE: 29-JUN-1990
 APPLICATION NUMBER: US 07/457,196
 FILING DATE: 29-DEC-1989
 APPLICATION NUMBER: US 07/390,901
 FILING DATE: 08-AUG-1989

ATTORNEY/AGENT INFORMATION:

NAME: Cserr, Luann
 REGISTRATION NUMBER: 31,822
 REFERENCE/DOCKET NUMBER: GI 5190
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (617)876-1170
 TELEFAX: (617)876-5851
 INFORMATION FOR SEQ ID NO: 104:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 1140 amino acids
 TYPE: amino acid
 TOPOLOGY: linear

```

; MOLECULE TYPE: protein
; SEQUENCE DESCRIPTION: SEQ ID NO: 104:
US-10-124-557-104.

Query Match          99.7%; Score 5811; DB 13; Length 1140;
Best Local Similarity 95.6%; Pred. No. 5,1e-143;
Matches 1090; Conservative 0; Mismatches 0; Indels 50; Gaps 1

QY      1 MAWKILPIYLLLLLSVFVIQVSSQDILSSCAGRCGEGYSRDATCNDYNCQHYMECCPDF 60
      |||
Db      1 MAWKILPIYLLLLLSVFVIQVSSQDILSSCAGRCGEGYSRDATCNDYNCQHYMECCPDF 60
      |||

QY     61 KVCVTAELSCKGRCPESFERGRCDDAQCCKYDKCCPDYESFCA----- 105
      |||
Db     61 KVCVTAELSCKGRCPESFERGRCDDAQCCKYDKCCPDYESFCAEVAHNPSPPPSSKKAP 120
      |||

QY     106 ----- -EEHVSVENQESSSSSSSSSSSSSIW 130
      |||
Db     121 PPSGASQIKSTTKRSPPNKKTKKVISESEITTEHVSVENQESSSSSSSSSSSSSIW 180
      |||

QY     131 KIKSKNSGAANRELOKLVKDNKNKTKKKPKPKPVVDGASGJLNDGDKVTTPTDST 190
      |||
Db     181 KIKSKNSGAANRELOKLVKDNKNKTKKKPKPKPVVDGASGJLNDGDKVTTPTDST 240
      |||

QY     191 TQHNKVSTSPKITTAKPINRPSLPNPSDTSKETSITVKNKETTIVETKETTINKQITSDG 250
      |||
Db     241 TQHNKVSTSPKITTAKPINRPSLPNPSDTSKETSITVKNKETTIVETKETTINKQITSDG 300
      |||

QY     251 KEKTTSAKETOSIEKTSAKDLAPTSKVLAKPTPKAETTKGPALTTPKBPTTTPEKAS 310
      |||
Db     301 KEKTTSAKETOSIEKTSAKDLAPTSKVLAKPTPKAETTKGPALTTPKBPTTTPEKAS 360
      |||

QY     311 TTPKBPRTTTIKSAPTTKEBAPTTKSAPTTKEBAPTTTKBAPTTTKBAPTTTKBAPTTTKBP 370
      |||
Db     361 TTPKBPRTTTIKSAPTTKEBAPTTKSAPTTKEBAPTTTKBAPTTTKBAPTTTKBAPTTTKBP 420
      |||

QY     371 APTTTKSAPTTPKBAPPTPKKAPPTPKBAPPTPKBAPPTPKBAPPTTKBAPPTTKBAPPTPK 430
      |||
Db     421 APTTTKSAPTTPKBAPPTPKKAPPTPKBAPPTPKBAPPTPKBAPPTPKBAPPTTKBAPPTPK 480
      |||

QY     431 EPAPTAPKKAPPTPKBAPTTPKBAPTTTKBPSPTTKBAPTTTKBAPTTTKBAPTTTKBAPTT 490
      |||
Db     481 EPAPTAPKKAPPTPKBAPTTPKBAPTTTKBPSPTTKBAPTTTKBAPTTTKBAPTTTKBAPTT 540
      |||

QY     491 TTKSAPTTPKBSPTTTKBAPTTPKBAPTTPKKAPTTPKBAPTTTKBAPTTTKBAPTTTKBP 550
      |||
Db     541 TTKSAPTTPKBSPTTTKBAPTTPKBAPTTPKKAPTTPKBAPTTTKBAPTTTKBAPTTTKBP 600
      |||

QY     551 APTAPKBPAPTTPKBAPTPTPKKLTPPTPKBAPTTPEKAPTTPELAPTTPEBPTPTT 610
      |||
Db     601 APTAPKBPAPTTPKBAPTPTPKKLTPPTPKBAPTTPEKAPTTPELAPTTPEBPTPTT 660
      |||

QY     611 PPEBAPTTPKAAAPNTPKBAPTTPKBAPTTPKBAPTTPKBAPTTPKETAAPTTPKGAPTTTLKEP 670
      |||
Db     661 PPEBAPTTPKAAAPNTPKBAPTTPKBAPTTPKBAPTTPKBAPTTPKETAAPTTPKGAPTTTLKEP 720
      |||

QY     671 APTTPKKAPAPKELAPTTTKBPTSTTSOKPAPTTPKGTAPTTPKBAPTTPKBAPTTPKBAPT 730
      |||
Db     721 APTTPKKAPAPKELAPTTTKBPTSTTSOKPAPTTPKGTAPTTPKBAPTTPKBAPTTPKBAPT 780
      |||

QY     731 TAPTTLKBAPPTPKKBPAPKELAPTTTKGPTSTTSOKPAPTTPKEAPTTPEKAPTTPEKAPTTPK 790
      |||
Db     781 TAPTTLKBAPPTPKKBPAPKELAPTTTKGPTSTTSOKPAPTTPKEAPTTPEKAPTTPEKAPTTPK 840
      |||

QY     791 KPAPTTPETPTPTTSEVSTPTTTKEPTTTIHKSPDESTPELSABPTPKALENSPKBPGVPT 850
      |||
Db     841 KPAPTTPETPTPTTSEVSTPTTTKEPTTTIHKSPDESTPELSABPTPKALENSPKBPGVPT 900
      |||

QY     851 TKTPAAATKPEMTTIAKQDITTBRLATPTPTTAAAPKMTKETAATTEKTTIESKITATTTCV 910
      |||
Db     901 TKTPAAATKPEMTTIAKQDITTBRLATPTPTTAAAPKMTKETAATTEKTTIESKITATTTCV 960
      |||

QY     911 TSTTTQDPTTFPKTITLTKTTTLAPKVVTTTKTITITTEIMNKPEETAKPKORATNKAITPK 970
      |||

```

| | | | |
|----|------|---|------|
| Db | 961 | TSVTTQTTDPFKITLTKTTTLLAPKVVTTTKITTTTTELNKPEETAKPKDRATNSKATTPK | 1020 |
| Qy | 971 | PQKPTKAPKKPTSTKKPKTMRVVRKPKXTTTPRKWTSTMPELNPTSRIA EAMLQTTTTPN | 1030 |
| Db | 1021 | PQKPTKAPKKPTSTKKPKTMRVVRKPKXTTTPRKWTSTMPELNPTSRIA EAMLQTTTTPN | 1080 |
| Qy | 1031 | QTPNSKLIVEYNPKSEDAGGAEGETPHMLLRPHVFMPEVTPMDYLPVPVQGIINPMLS | 1090 |
| Db | 1081 | QTPNSKLIVEYNPKSEDAGGAEGETPHMLLRPHVFMPEVTPMDYLPVPVQGIINPMLS | 1140 |

RESULT 3
 US-09-802-207-30
 ; Sequence 30, Application US/09802207
 ; Publication No. US20020086824A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Warman, Matthew
 ; APPLICANT: Carpten, John
 ; APPLICANT: Trent, Jeffrey
 ; APPLICANT: Marcelino, Jose
 ; TITLE OF INVENTION: Novel Methods and Reagents for the Treatment of Osteoarthritis
 ; FILE REFERENCE: Case 06212
 ; CURRENT APPLICATION NUMBER: US/09/802,207
 ; CURRENT FILING DATE: 2001-08-29
 ; PRIOR APPLICATION NUMBER: 09/619,175
 ; PRIOR FILING DATE: 2000-07-19
 ; PRIOR APPLICATION NUMBER: 60/145,328
 ; PRIOR FILING DATE: 1999-07-23
 ; NUMBER OF SEQ ID NOS: 30
 ; SOFTWARE: PatentIn version 3.0
 ; SEQ ID NO 30
 ; LENGTH: 1404
 ; TYPE: PRT
 ; ORGANISM: Homo sapiens
 US-09-802-207-30

| Query Match | 99.7% | Score 5811 | DB 9 | Length 1404 |
|-----------------------|----------------|--|-----------|-------------|
| Best Local Similarity | 95.6% | Pred. No. 6.5e-143 | | |
| Matches 1090 | Conservative 0 | Mismatches 0 | Indels 50 | Gaps 1 |
| Qy | 1 | MAWKTLPIYLLLLSVFVIOQVSSQDLSSSCAGRCGEGYSRDATCNCDCYNQHYMECCPDF | 60 | |
| Db | 1 | MAWKTLPIYLLLLSVFVIOQVSSQDLSSSCAGRCGEGYSRDATCNCDCYNQHYMECCPDF | 60 | |
| Qy | 61 | KRVCTAELSCKGRCFESFERGECDDAQCKKYDKCCPDYESFCA----- | 105 | |
| Db | 61 | KRVCTAELSCKGRCFESFERGECDDAQCKKYDKCCPDYESFCAEYHNPTSPSSKKAP | 120 | |
| Qy | 106 | -----EEHVSSENQESSSSSSSSSSSSSSSSSSSSSIW | 130 | |
| Db | 121 | PPSGASQTIKSTYKRSPPKPNKKTKVIRSEETIEHVSSENQESSSSSSSSSSSSSSSIW | 180 | |
| Qy | 131 | KIKSSKNSAANRELQKLVKQDNKKNTKKKPTKPPVVDAGSLONGDFKVTTPDTST | 190 | |
| Db | 181 | KIKSSKNSAANRELQKLVKQDNKKNTKKKPTKPPVVDAGSLONGDFKVTTPDTST | 240 | |
| Qy | 191 | TOHNKUSTSKPIITTAKEINRPSPLPNNSDTSKETSLTVNKEITTVETKETTITNKQTSIDG | 250 | |
| Db | 241 | TOHNKUSTSKPIITTAKEINRPSPLPNNSDTSKETSLTVNKEITTVETKETTITNKQTSIDG | 300 | |
| Qy | 251 | KEKITSAKETQSIKTSKAKOLAPTSKVLAKPTPKAETTTKGPAITTPKEPTPTTPKEPAS | 310 | |
| Db | 301 | KEKITSAKETQSIKTSKAKOLAPTSKVLAKPTPKAETTTKGPAITTPKEPTPTTPKEPAS | 360 | |
| Qy | 311 | TTTKEPTPTTIKSAPTTPKBPATTTIKSAPTTPKBPATTTKEBPATTTKEBPATTTTKEP | 370 | |
| Db | 361 | TTTKEPTPTTIKSAPTTPKBPATTTIKSAPTTPKBPATTTTKEBPATTTTKEBPATTTTKEP | 420 | |
| Qy | 371 | APTITTKGAPTTKBPAPTTTKKBPATTPKBPATTPKPEPTPTTKEBPATTTKEBPATTPK | 430 | |
| Db | 421 | APTITTKGAPTTKBPATTPKBPATTPKBPATTPKPEPTPTTKEBPATTTKEBPATTPK | 480 | |

RESULT 5
US-10-124-557-2
; Sequence 2, Application US/10124557
; Publication No. US20020137894A1
; GENERAL INFORMATION:
; APPLICANT: Turner, Katherine
; Clark, Stephen C.
; Jacobs, Kenneth
; Hewick, Rodney M.
; Gesner, Thomas G.
; TITLE OF INVENTION: Megakaryocyte Stimulating Factors
; NUMBER OF SEQUENCES: 143
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genetics Institute, Inc.
; STREET: 87 CambridgePark Drive
; CITY: Cambridge
; STATE: Massachusetts
; COUNTRY: U.S.A.
; ZIP: 02140
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/10/124,557
; FILING DATE: 16-Apr-2002
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/643,502
; FILING DATE: 18-JAN-1991
; APPLICATION NUMBER: US 07/546,114
; FILING DATE: 29-JUN-1990
; APPLICATION NUMBER: US 07/457,196
; FILING DATE: 29-DEC-1989
; APPLICATION NUMBER: US 07/390,901
; FILING DATE: 08-AUG-1989
; ATTORNEY/AGENT INFORMATION:
; NAME: Cserr, Luann
; REGISTRATION NUMBER: 31,922
; REFERENCE/DOCKET NUMBER: GI 5190
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617)876-1170
; TELEFAX: (617)876-5851
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1404 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; SEQUENCE DESCRIPTION: SEQ ID NO: 2:
US-10-124-557-2
Query Match 99.7%; Score 5811; DB 13; Length 1404;
Best Local Similarity 95.6%; Pred. No. 6.5e-143;
Matches 1090; Conservative 0; Mismatches 0; Indels 50; Gaps 1;
QY 1 MAWKTLPIYLLLLSVFVIQVSSQDLSSCAGRCGEGYGRDATCNCYNCQHYMECCPDF 60
DB 1 MAWKTLPIYLLLLSVFVIQVSSQDLSSCAGRCGEGYGRDATCNCYNCQHYMECCPDF 60
QY 61 KRVTAEALCKGRCFESFGRGECDDAOCKYDKCCPDYESFCA----- 105
DB 61 KRVTAEALCKGRCFESFGRGECDDAOCKYDKCCPDYESFCAVHNPTSPPSKXAP 120
QY 106 -----EHSVSNQSSSSSSSSSSSSSSSIW 130
DB 121 PPSGASQTIKSTTKRSPKPNKKTKVIESIEIIEHSVSNQSSSSSSSSSSSIW 180
QY 131 KIKSSKNSAANRELQKKLVKDNKNKRTKKKPTPKPPVVDEAGSLDNGDFKVTPTDST 190

RESULT 6
US-10-124-557-62
; Sequence 62, Application US/10124557
; Publication No. US20020137894A1
; GENERAL INFORMATION:
; APPLICANT: Turner, Katherine
; Clark, Stephen C.
; Jacobs, Kenneth
; Hewick, Rodney M.
; Gesner, Thomas G.

TITLE OF INVENTION: Megakaryocyte Stimulating Factors
 NUMBER OF SEQUENCES: 143
 CORRESPONDENCE ADDRESS:

ADDRESSEE: Genetics Institute, Inc.
 STREET: 87 CambridgePark Drive
 CITY: Cambridge
 STATE: Massachusetts
 COUNTRY: U.S.A.
 ZIP: 02140

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: PatentIn Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/10/124,557

FILING DATE: 16-Apr-2002

CLASSIFICATION: <Unknown>

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 07/643,502

FILING DATE: 18-JAN-1991

APPLICATION NUMBER: US 07/546,114

FILING DATE: 29-JUN-1990

APPLICATION NUMBER: US 07/457,196

FILING DATE: 29-DEC-1989

APPLICATION NUMBER: US 07/390,901

FILING DATE: 08-AUG-1989

ATTORNEY/AGENT INFORMATION:

NAME: Cserr, Luann

REGISTRATION NUMBER: 31,822

REFERENCE/DOCKET NUMBER: GI 5190

TELECOMMUNICATION INFORMATION:

TELEPHONE: (617)876-1170

TELEFAX: (617)876-5851

INFORMATION FOR SEQ ID NO: 62:

SEQUENCE CHARACTERISTICS:

LENGTH: 1404 amino acids

TYPE: amino acid

TOPOLOGY: linear

MOLECULE TYPE: protein

SEQUENCE DESCRIPTION: SEQ ID NO: 62:

US-10-124-557-62
 Query Match 99.7%; Score 5811; DB 13; Length 1404;
 Best Local Similarity 95.6%; Pred. No. 6.5e-143;
 Matches 1090; Conservative 0; Mismatches 0; Indels 50; Gaps 1;

QY 1 MAWKTLPIYLLILLVSVFVIQVSSQDLSSCAGRCGEGYSRDATCNCYDNCQHYMECCPDF 60
 DB 1 MAWKTLPIYLLILLVSVFVIQVSSQDLSSCAGRCGEGYSRDATCNCYDNCQHYMECCPDF 60
 QY 61 KRVTCTAEISCKGRCPESFERGECDCDAQCKYDKCCPDYBSPCA----- 105
 DB 61 KRVTCTAEISCKGRCPESFERGECDCDAQCKYDKCCPDYBSPCAVHNFTSPSSKKAP 120
 QY 106 -----EEHVSSENQESSSSSSSSSSSTIW 130
 DB 121 PPSGASQIKSTTKSPKPPNKKTKVIESEITEHVSSENQESSSSSSSSSSSTIW 180
 QY 131 KIKSKNSAANRELQKKLVKNDKNKNTKKTPTKPPVVDAGSGLDNGDFKVTPTST 190
 DB 181 KIKSKNSAANRELQKKLVKNDKNKNTKKTPTKPPVVDAGSGLDNGDFKVTPTST 240
 QY 191 TQHNKVSTSPKITTAKPINRPSLPNPSDTSKESLTVNKETTVETKTTTNKQTSIDG 250
 DB 241 TQHNKVSTSPKITTAKPINRPSLPNPSDTSKESLTVNKETTVETKTTTNKQTSIDG 300
 QY 251 KEKTSKETSISKTSKADLAPTSKVLAETPKAETTTKGPALTTPKEPTTPPKEPAS 310
 DB 301 KEKTSKETSISKTSKADLAPTSKVLAETPKAETTTKGPALTTPKEPTTPPKEPAS 360
 QY 311 TTPKEPTTTIKSAPTTPKEPATTTKSAPTTPKEPATTTKEPAPTTKEPAPTTKEPAPTTTKEP 370

361 TTPKEPTTTIKSAPTTPKEPATTTKSAPTTPKEPATTTKEPAPTTTKEPAPTTTKEPAPTTTKEP 420
 QY 371 APTTTKSAPTTPKEPATTTPKEPATTTPKEPATTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTT 430
 DB 421 APTTTKSAPTTPKEPATTTPKEPATTTPKEPATTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTT 480
 QY 431 EPAPTAPKPKAPATTTPKEPATTTPKEPATTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPT 490
 DB 481 EPAPTAPKPKAPATTTPKEPATTTPKEPATTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPT 540
 QY 491 TTKSAPTTPKESPSTTTKEPAPTTPKEPATTTPKEPATTTKPKAPATTTPKEPAPTTTKEPAPTTTKEPAPT 550
 DB 541 TTKSAPTTPKESPSTTTKEPAPTTPKEPATTTPKEPATTTKPKAPATTTPKEPAPTTTKEPAPTTTKEPAPT 600
 QY 551 APTAPKPEAPTTPKETAPTTPKKLPTTTBEKLAPTTPEKAPATTTPEKAPATTTPEELAPTTPEEPTPTT 610
 DB 601 APTAPKPEAPTTPKETAPTTPKKLPTTTBEKLAPTTPEKAPATTTPEKAPATTTPEELAPTTPEEPTPTT 660
 QY 611 PEEPAPTTPKAAAPNTPKEPATTTPKEPATTTPKEPATTTPKETAPTTPKETAPTTPKGTAPTTTLKEP 670
 DB 661 PEEPAPTTPKAAAPNTPKEPATTTPKEPATTTPKEPATTTPKETAPTTPKETAPTTPKGTAPTTTLKEP 720
 QY 671 APTTPKPKAPKELAPTTTKEPTSTTSDKAPATTTKGTAPTTTKEPAPATTTKEPAPATTTKEPAPATTKG 730
 DB 721 APTTPKPKAPKELAPTTTKEPTSTTSDKAPATTTKGTAPTTTKEPAPATTTKEPAPATTTKEPAPATTKG 780
 QY 731 TAPTTTLKEPAPTTPKKAPKELAPTTTKEPTSTTSDKAPATTTKGTAPTTTKEPAPATTTKEPAPATTTKEPAPT 790
 DB 781 TAPTTTLKEPAPTTPKKAPKELAPTTTKEPTSTTSDKAPATTTKGTAPTTTKEPAPATTTKEPAPATTTKEPAPT 840
 QY 791 KPAPTTPEPTTPEPTTSEVSTPTTTKEPTTIHKSPDSTPSELSAEPSTKALENSPKSPGVPT 850
 DB 841 KPAPTTPEPTTPEPTTSEVSTPTTTKEPTTIHKSPDSTPSELSAEPSTKALENSPKSPGVPT 900
 QY 851 TKTPAATKPEMTTAKDKTTERDLRTTPETTTAAKPTWKETAATTTKEKTTESKITATTTQV 910
 DB 901 TKTPAATKPEMTTAKDKTTERDLRTTPETTTAAKPTWKETAATTTKEKTTESKITATTTQV 960
 QY 911 TSTTTQDTPPFKITTLLKTTLAPKVTITTTTINMKPEETAKPKDRATNSKATTPK 970
 DB 961 TSTTTQDTPPFKITTLLKTTLAPKVTITTTTINMKPEETAKPKDRATNSKATTPK 1020
 QY 971 POKPTKAPKPTSTKKEKTPRVRKPTTPTPKVTSTMPBLNPTSRABAMLOTTTTPN 1030
 DB 1021 POKPTKAPKPTSTKKEKTPRVRKPTTPTPKVTSTMPBLNPTSRABAMLOTTTTPN 1080
 QY 1031 QTPNSKLVNPKSEDAAGAEGETPHMLLRPHVFPVFTPDMDYLPRVFNQGIINPMLS 1090
 DB 1081 QTPNSKLVNPKSEDAAGAEGETPHMLLRPHVFPVFTPDMDYLPRVFNQGIINPMLS 1140

RESULT 7

US-10-124-557-40

; Sequence 40, Application US/10124557

; Publication No. US20020137894A1

; GENERAL INFORMATION:

; APPLICANT: Turner, Katherine

; Jacobs, Kenneth

; Hawick, Rodney M.

; Gesner, Thomas G.

; TITLE OF INVENTION: Megakaryocyte Stimulating Factors

; NUMBER OF SEQUENCES: 143

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Genetics Institute, Inc.

; STREET: 87 CambridgePark Drive

; CITY: Cambridge

; STATE: Massachusetts

; COUNTRY: U.S.A.

; ZIP: 02140

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/10/124,557

FILING DATE: 16-Apr-2002

CLASSIFICATION: <unknown>

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 07/643,502

FILING DATE: 18-JAN-1991

APPLICATION NUMBER: US 07/546,114

FILING DATE: 29-JUN-1990

APPLICATION NUMBER: US 07/457,196

FILING DATE: 29-DEC-1989

APPLICATION NUMBER: US 07/390,901

FILING DATE: 08-AUG-1989

ATTORNEY/AGENT INFORMATION:

NAME: Cserr, Juann

REGISTRATION NUMBER: 31,822

REFERENCE/DOCKET NUMBER: GI 5190

TELECOMMUNICATION INFORMATION:

TELEPHONE: (617)876-1170

TELEFAX: (617)876-5851

INFORMATION FOR SEQ ID NO: 40:

SEQUENCE CHARACTERISTICS:

LENGTH: 1361 amino acids

TYPE: amino acid

TOPOLOGY: linear

MOLECULE TYPE: protein

SEQUENCE DESCRIPTION: SEQ ID NO: 40:

US-10-124-557-40

Query Match 96.9%; Score 5643.3; DB 13; Length 1361;

Best Local Similarity 96.3%; Pred. No. 1.4e-138;

Matches 1056; Conservative 26; Indels 7; Gaps 1;

1 MAWKLPYVILLVLLVVFVQVSSQDLSSCAGCGEGYSRDATCNCYNCOHYMECCPDF 60

1 MAWKLPYVILLVLLVVFVQVSSQDLSSCAGCGEGYSRDATCNCYNCOHYMECCPDF 60

61 KRVTAEKCKGRCFESFERGECDDAOKKYDKCCPDYEFCAEHSVSSENQESSSS 120

61 KRVTAEKCKGRCFESFERGECDDAOKKYDKCCPDYEFCAEHSVSSENQESSSS 120

121 SSSSSSTTWIKSSXNSAANRELQK-----KVONKQNRKTKKTPPVVDKAG 173

121 PPSGASQTIKSTTRGPKPPNKKTKVIESEBITEVKNKNRTKKTPPVVDKAG 180

174 SGLDNGDFKVTTPDTSTOHNKVSTSPKITTAKPINRPSLPNSDTSKETSITVNNKETT 233

181 SGLDNGDFKVTTPDTSTOHNKVSTSPKITTAKPINRPSLPNSDTSKETSITVNNKETT 240

234 VETKETTITNKQSTDGKETTSAKETQSIKTSKADLAPTSKVLAKTPKAEITTKGPA 293

241 VETKETTITNKQSTDGKETTSAKETQSIKTSKADLAPTSKVLAKTPKAEITTKGPA 300

294 LTTPKETPTTPKEPASTTPKETPTTIKSAPTTPKEPAPTTTKSAPTTPKEPAPTTTKE 353

301 LTTPKETPTTPKEPASTTPKETPTTIKSAPTTPKEPAPTTTKSAPTTPKEPAPTTTKE 360

354 PAPTTPKEPAPTTTKSAPTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPTPTT 413

361 PAPTTPKEPAPTTTKSAPTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPTPTT 420

414 PKEPAPTTKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPA 473

421 PKEPAPTTKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPA 480

474 PTTTKSAPTTKKEPAPTTTKSAPTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTT 533

481 PTTTKSAPTTKKEPAPTTTKSAPTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTT 540

534 PAPTTPKEPAPTTTKKAP 593

541 PAPTTPKEPAPTTTKKAP 600

594 TPEELAPTTPEBPTPTTPEEPAPTTPKAAAPNTPEEPAPTTPEEPAPTTPEEPAPTTPE 653

601 TPEELAPTTPEBPTPTTPEEPAPTTPKAAAPNTPEEPAPTTPEEPAPTTPEEPAPTTPE 660

654 TAPTTPKGTAPTTLKKEPAPTTPKKAPKELAPTTTKEPTSTTSKAPAPTTPKGTAPTTK 713

661 TAPTTPKGTAPTTLKKEPAPTTPKKAPKELAPTTTKEPTSTTSKAPAPTTPKGTAPTTK 720

714 EPAPTTPKSPAPTTPKGTAPTTLKKEPAPTTPKKAPKELAPTTTKEPTSTTSKAPAPTT 773

721 EPAPTTPKSPAPTTPKGTAPTTLKKEPAPTTPKKAPKELAPTTTKEPTSTTSKAPAPTT 780

774 KETAPTTPEEPAPTTPKKAPAPTTPEPTTSEVSTPTTKEPTTTHKSPDSESTPELSAE 833

781 KETAPTTPEEPAPTTPKKAPAPTTPEPTTSEVSTPTTKEPTTTHKSPDSESTPELSAE 840

834 PTPKALENSPKGVPVTTTTPAATKPEMTTAKDKTERDLRTTPTTTAAAPQWTKETAT 893

841 PTPKALENSPKGVPVTTTTPAATKPEMTTAKDKTERDLRTTPTTTAAAPQWTKETAT 900

894 TTEKTTESKITATTQVSTTTTQDTTTPFKITLLKTTTLAPKVTITTKITTTIMNKPEE 953

901 TTEKTTESKITATTQVSTTTTQDTTTPFKITLLKTTTLAPKVTITTKITTTIMNKPEE 960

954 TAKPKDRATNSKATTPKPKQPTKAPKPTSTKPKTPRVRKPKTTPRKMSTWPELN 1013

961 TAKPKDRATNSKATTPKPKQPTKAPKPTSTKPKTPRVRKPKTTPRKMSTWPELN 1020

1014 PTSRIAEAMLQTTTRPNCTPNSKLVNPKSEDAAGAGETPHMLLRPHVFMPEVTPDMD 1073

1021 PTSRIAEAMLQTTTRPNCTPNSKLVNPKSEDAAGAGETPHMLLRPHVFMPEVTPDMD 1080

1074 YLPRVNPQGIINPMLS 1090

1081 YLPRVNPQGIINPMLS 1097

RESULT 8

US-10-124-557-42

Sequence 42, Application US/10124557

Publication No. US20020137894A1

GENERAL INFORMATION:

APPLICANT: Turner, Katherine

Clark, Stephen C.

Jacobs, Kenneth

Hewick, Rodney M.

Gesner, Thomas G.

TITLE OF INVENTION: Megakaryocyte Stimulating Factors

NUMBER OF SEQUENCES: 143

CORRESPONDENCE ADDRESS:

ADDRESSEE: Genetics Institute, Inc.

STREET: 87 CambridgePark Drive

CITY: Cambridge

STATE: Massachusetts

COUNTRY: U.S.A.

ZIP: 02140

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patent In Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/10/124,557

FILING DATE: 16-Apr-2002

CLASSIFICATION: <Unknown>

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 07/643,502

FILING DATE: 18-JAN-1991

APPLICATION NUMBER: US 07/546,114

FILING DATE: 29-JUN-1990

APPLICATION NUMBER: US 07/457,196

```

; FILING DATE: 29-DEC-1989
; APPLICATION NUMBER: US 07/390,901
; FILING DATE: 08-AUG-1989
; ATTORNEY/AGENT INFORMATION:
; NAME: Cseerz, Luann
; REGISTRATION NUMBER: 31,822
; REFERENCE/DOCKET NUMBER: GI 5190
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617)876-1170
; TELEFAX: (617)876-5851
; INFORMATION FOR SEQ ID NO: 42:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1311 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; SEQUENCE DESCRIPTION: SEQ ID NO: 42:
US-10-124-557-42

Query Match          96.3%; Score 5608.7; DB 13; Length 1311;
Best Local Similarity 96.1%; Pred. No. 1.1e-137;
Matches 1047; Conservative 0; Mismatches 0; Indels 43; Gaps 1;

QY 1 MAWKTLPIYLLLLLSVFVIQVSSODLSACGRCGEGYSRDATCNCYDNCQHYMECCPDF 60
DB 1 MAWKTLPIYLLLLLSVFVIQVSSODLSACGRCGEGYSRDATCNCYDNCQHYMECCPDF 60
QY 61 KVVCTAELSCGRCFESPERGECDCDAQCKKYDKCCPDYDFCAEHSVSENQESSSS 120
DB 61 KVVCTAELSCGRCFESPERGECDCDAQCKKYDKCCPDYDFCAE----- 106
QY 121 SSSSSSSITWIKSSKNSAANRELKLVKDNKQVTKKPTPKPPVVDAGSLDNGD 180
DB 107 -----VKDNKKNRTKKKPTPKPPVVDAGSLDNGD 137
QY 181 FKVTPTDSTTQHNKVSPTKITTAKPINRPSLPNSDTSKETSLSLVNKETTVEKETT 240
DB 138 FKVTPTDSTTQHNKVSPTKITTAKPINRPSLPNSDTSKETSLSLVNKETTVEKETT 197
QY 241 TNNKQTSDDGKEKITSKETOSIEKTSKOLAPTSKVLAKPTPKAETTTKGPALTTPKEP 300
DB 198 TNNKQTSDDGKEKITSKETOSIEKTSKOLAPTSKVLAKPTPKAETTTKGPALTTPKEP 257
QY 301 TPTTPKEPASTTPKEPTPTTIKSAPTTPKEPAPTTTKSAPTTPKEPAPTTTKPAPTT 360
DB 258 TPTTPKEPASTTPKEPTPTTIKSAPTTPKEPAPTTTKSAPTTPKEPAPTTTKPAPTT 317
QY 361 EPAPTTKEPAPTTTKSAPTTPKEPAPTTTKPAPTTTKPAPTTTKPAPTTTKPAPTT 420
DB 318 EPAPTTKEPAPTTTKSAPTTPKEPAPTTTKPAPTTTKPAPTTTKPAPTTTKPAPTT 377
QY 421 TKPAPTTKEPAPTTAKPAPTTTKPAPTTTKPAPTTTKPAPTTTKPAPTTTKPAPTT 480
DB 378 TKPAPTTKEPAPTTAKPAPTTTKPAPTTTKPAPTTTKPAPTTTKPAPTTTKPAPTT 437
QY 481 PTTTKEPAPTTTKSAPTTPKEPAPTTTKPAPTTTKPAPTTTKPAPTTTKPAPTTTPK 540
DB 438 PTTTKEPAPTTTKSAPTTPKEPAPTTTKPAPTTTKPAPTTTKPAPTTTKPAPTTTPK 497
QY 541 EPAPTTTKKAPAPTKPAPTTTKPAPTTTKPAPTTTKPAPTTTKPAPTTTKPAPTTPE 600
DB 498 EPAPTTTKKAPAPTKPAPTTTKPAPTTTKPAPTTTKPAPTTTKPAPTTTKPAPTTPE 557
QY 601 TTPEEPTPTTPEPAPTTTKAANPTPKPAPTTTKPAPTTTKPAPTTTKPAPTTTKPAP 660
DB 558 TTPEEPTPTTPEPAPTTTKAANPTPKPAPTTTKPAPTTTKPAPTTTKPAPTTTKPAP 617
QY 661 GTAPTTLKEPAPTTPKAPKELAPTTTKPTSTSDKAPTTTKGAPTTPKETAPTTP 720
DB 618 GTAPTTLKEPAPTTPKAPKELAPTTTKPTSTSDKAPTTTKGAPTTPKETAPTTP 677
QY 721 KEPAPTTPKGAPTTLKEPAPTTPKAPKELAPTTTKGPTSTSDKAPTTTKGAPTTP 780

```

```

DB 678 KEPAPTTPKGAPTTLKEPAPTTPKAPKELAPTTTKGPTSTSDKAPTTTKETAPTT 737
QY 781 PKPAPTTPKKAPAPTTTETPTTSEVSTTTTKEPTTIHKSPDESTPELSAETPKALE 840
DB 738 PKPAPTTPKKAPAPTTTETPTTSEVSTTTTKEPTTIHKSPDESTPELSAETPKALE 797
QY 841 NSPKPGVPTTKTAPKAPKEMTTAKDITERRDLRTTPTTAAPOKTKETATTTKTE 900
DB 798 NSPKPGVPTTKTAPKAPKEMTTAKDITERRDLRTTPTTAAPOKTKETATTTKTE 857
QY 901 SKITATTQVSTTTTQDTPFKITTLKTTTLAPKVTTTKTITTTIMNKPETAKPKDR 960
DB 858 SKITATTQVSTTTTQDTPFKITTLKTTTLAPKVTTTKTITTTIMNKPETAKPKDR 917
QY 961 ATNSKATTPKQPTKAPKPTSTKPKMVRVKPTTTPRKMSTWPELNPTSRIAE 1020
DB 918 ATNSKATTPKQPTKAPKPTSTKPKMVRVKPTTTPRKMSTWPELNPTSRIAE 977
QY 1021 AMLQTTTRNQTNSKLVNPKSEDAAGAGETPHMLLRPHVFMDEVTPDMDYLPVRVN 1080
DB 978 AMLQTTTRNQTNSKLVNPKSEDAAGAGETPHMLLRPHVFMDEVTPDMDYLPVRVN 1037
QY 1081 QGIIINPMLS 1090
DB 1038 QGIIINPMLS 1047

```

RESULT 9

```

US-10-124-557-50
; Sequence 50, Application US/10124557
; Publication No. US2002037894A1
; GENERAL INFORMATION:
; APPLICANT: Turner, Katherine
;           Clark, Stephen C.
;           Jacobs, Kenneth
;           Hewick, Rodney M.
;           Gesner, Thomas G.
; TITLE OF INVENTION: Megakaryocyte Stimulating Factors
; NUMBER OF SEQUENCES: 143
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genetics Institute, Inc.
; STREET: 87 Cambridgepark Drive
; CITY: Cambridge
; STATE: Massachusetts
; COUNTRY: U.S.A.
; ZIP: 02140
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent in Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/10/124,557
; FILING DATE: 16-Apr-2002
; CLASSIFICATION: <unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/643,502
; FILING DATE: 18-JAN-1991
; APPLICATION NUMBER: US 07/546,114
; FILING DATE: 29-JUN-1990
; APPLICATION NUMBER: US 07/457,196
; FILING DATE: 29-DEC-1989
; APPLICATION NUMBER: US 07/390,901
; FILING DATE: 08-AUG-1989
; ATTORNEY/AGENT INFORMATION:
; NAME: Cseerz, Luann
; REGISTRATION NUMBER: 31,822
; REFERENCE/DOCKET NUMBER: GI 5190
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617)876-1170
; TELEFAX: (617)876-5851
; INFORMATION FOR SEQ ID NO: 50:
; SEQUENCE CHARACTERISTICS:

```

```

;
; LENGTH: 1314 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; SEQUENCE DESCRIPTION: SEQ ID NO: 50:
US-10-124-557-50

Query Match          95.6%; Score 5571; DB 13; Length 1314;
Best Local Similarity 96.3%; Pred. No. 1e-136;
Matches 1050; Conservative 0; Mismatches 0; Indels 40; Gaps 1;

QY 1 MAWKTLPIYLLLSLVFVIQQVSSQDLSSCAGRCGEGYSRDATCNCYHMECCPDF 60
DB 1 MAWKTLPIYLLLSLVFVIQQVSSQDLSSCAGRCGEGYSRDATCNCYHMECCPDF 60
QY 61 KXVCTAELSCGRCPESPERGECDCDAQCKYDKCCPDYSEFCAEHSVSENOESSSS 120
DB 61 KXVCTA-----EHSVSENOESSSS 80
QY 121 SSSSSSTIWKTKSKNGAANRELQKLVKDNKNKRTKKKPTKPPVVDVAGSLDNGD 180
DB 81 SSSSSSTIWKTKSKNGAANRELQKLVKDNKNKRTKKKPTKPPVVDVAGSLDNGD 140
QY 181 FKVTTPTDSTTOHNVKVSPIKTTAKPINRPSLPNSDTSKETSLSLVNKTIVETKETT 240
DB 141 FKVTTPTDSTTOHNVKVSPIKTTAKPINRPSLPNSDTSKETSLSLVNKTIVETKETT 200
QY 241 ITNKQTSIDGKEKTTSAKETQSIKTSKADLAPTSKVLAKPTPKAETTTKGPALTTKPEP 300
DB 201 ITNKQTSIDGKEKTTSAKETQSIKTSKADLAPTSKVLAKPTPKAETTTKGPALTTKPEP 260
QY 301 TPTTKEPASTTPKEPTPTTIKASATTPKEPAPTTTKSAPTTKEPAPTTTKEPAPTTPK 360
DB 261 TPTTKEPASTTPKEPTPTTIKASATTPKEPAPTTTKSAPTTKEPAPTTTKEPAPTTPK 320
QY 361 EPAPTTTKEPAPTTTKSAPTTKEPAPTTPKPAPTTTKEPAPTTTKEPAPTTTKEPAPT 420
DB 321 EPAPTTTKEPAPTTTKSAPTTKEPAPTTPKPAPTTTKEPAPTTTKEPAPTTTKEPAPT 380
QY 421 TKEPAPTTTKEPAPTAAPKAPATTPKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKE 480
DB 381 TKEPAPTTTKEPAPTAAPKAPATTPKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKE 440
QY 481 PTTTKEPAPTTTKSAPTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKE 540
DB 441 PTTTKEPAPTTTKSAPTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKE 500
QY 541 EPAPTTTKEPAPTAAPKAPATTPKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPT 600
DB 501 EPAPTTTKEPAPTAAPKAPATTPKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPT 560
QY 601 TTPKEPTTTPKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKE 660
DB 561 TTPKEPTTTPKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKE 620
QY 661 GTAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPT 720
DB 621 GTAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPT 680
QY 721 KEAPPTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPT 780
DB 681 KEAPPTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPT 740
QY 781 PKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPT 840
DB 741 PKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPT 800
QY 841 NSPKFVGPTTKTAAKPEMTTAAKDKTERDLRTPTTPTTAAKPMWTKETATTEKTE 900
DB 801 NSPKFVGPTTKTAAKPEMTTAAKDKTERDLRTPTTPTTAAKPMWTKETATTEKTE 860
QY 901 SKITATTQVTSSTTQDTTTPFKITTLTKTTTLAPKVTTKTKITTTTTEIMNKPEETAKPKDR 960

```

```

DB 961 SKITATTQVTSSTTQDTTTPFKITTLTKTTTLAPKVTTKTKITTTTTEIMNKPEETAKPKDR 920
QY 961 ATNSKATTPKQPKPTKAPKPKPTSTKKPKTMPRVRKPKTTTTPRKMSTMTPELNPTSRIAE 1020
DB 921 ATNSKATTPKQPKPTKAPKPKPTSTKKPKTMPRVRKPKTTTTPRKMSTMTPELNPTSRIAE 980
QY 1021 AMLQTTTRNQTPNSKLVNPKSEDAAGAGETPHMLLRPHVFMPEVTPDMDYLPRVPN 1080
DB 981 AMLQTTTRNQTPNSKLVNPKSEDAAGAGETPHMLLRPHVFMPEVTPDMDYLPRVPN 1040
QY 1081 QGIIINPMLS 1090
DB 1041 QGIIINPMLS 1050

RESULT 10
US-10-124-557-58
; Sequence 58, Application US/10124557
; Publication No US20020137894A1
; GENERAL INFORMATION:
; APPLICANT: Turner, Katherine
; Clark, Stephen C.
; Jacobs, Kenneth
; Hewick, Rodney M.
; Gesner, Thomas G.
; TITLE OF INVENTION: Megakaryocyte Stimulating Factors
; NUMBER OF SEQUENCES: 143
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genetics Institute, Inc.
; STREET: 87 Cambridgepark Drive
; CITY: Cambridge
; STATE: Massachusetts
; COUNTRY: U.S.A.
; ZIP: 02140
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/10/124,557
; FILING DATE: 16-Apr-2002
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/643,502
; FILING DATE: 18-JAN-1991
; APPLICATION NUMBER: US 07/546,114
; FILING DATE: 29-JUN-1990
; APPLICATION NUMBER: US 07/457,196
; FILING DATE: 29-DEC-1989
; APPLICATION NUMBER: US 07/390,901
; FILING DATE: 08-AUG-1989
; ATTORNEY/AGENT INFORMATION:
; NAME: Cseir, Luann
; REGISTRATION NUMBER: 31,822
; REFERENCE/DOCKET NUMBER: GI 5190
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617) 876-1170
; TELEFAX: (617) 876-5851
; INFORMATION FOR SEQ ID NO: 58:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1049 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; SEQUENCE DESCRIPTION: SEQ ID NO: 58:
US-10-124-557-58

Query Match          95.5%; Score 5561.9; DB 13; Length 1049;
Best Local Similarity 96.2%; Pred. No. 1.3e-136;
Matches 1049; Conservative 0; Mismatches 0; Indels 41; Gaps 1;

QY 1 MAWKTLPIYLLLSLVFVIQQVSSQDLSSCAGRCGEGYSRDATCNCYHMECCPDF 60

```

Db 1 MAWKTLPIYLLLSLVFVIQQVSSQ----- 25

QY 61 KRVCYTAELSCGRCFESFERGECDCDAQCKYDKCCPDYBSFCAEHSVSENQESSSS 120

Db 26 -----ELSCGRCFESFERGECDCDAQCKYDKCCPDYBSFCAEHSVSENQESSSS 79

QY 121 SSSSSSTIWKIKSSKNSAANRELOKLVKDNKKNRTKKKPTPKPPVVDAGSLDNGD 180

Db 80 SSSSSSTIWKIKSSKNSAANRELOKLVKDNKKNRTKKKPTPKPPVVDAGSLDNGD 139

QY 181 FKVTTPDSTTCHNKVSTSPKLTAKPINRPSLPNSDTSKETSLSLVNKETTIVETKETT 240

Db 140 FKVTTPDSTTCHNKVSTSPKLTAKPINRPSLPNSDTSKETSLSLVNKETTIVETKETT 199

QY 241 TTNKOTSDGKEKTTSAKETQSIKTSADLAPTSKVILAKPTPKAETTTKGPALTTTPKEP 300

Db 200 TTNKOTSDGKEKTTSAKETQSIKTSADLAPTSKVILAKPTPKAETTTKGPALTTTPKEP 259

QY 301 TPTTPKEPASTTPKEPTTTIKSAETPKAPATTTKSAPTTPKEPAPTTPKEPAPTTPK 360

Db 260 TPTTPKEPASTTPKEPTTTIKSAETPKAPATTTKSAPTTPKEPAPTTPKEPAPTTPK 319

QY 361 EPAPTTPKEPAPTTPKSAPTTPKEPAPTTPKAPATTPKEPAPTTPKEPAPTTPKEPAPT 420

Db 320 EPAPTTPKEPAPTTPKSAPTTPKEPAPTTPKAPATTPKEPAPTTPKEPAPTTPKEPAPT 379

QY 421 TKEPAPTTPKEPAPTAPKAPATTPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKSA 480

Db 380 TKEPAPTTPKEPAPTAPKAPATTPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKSA 439

QY 481 PTTTPKEPAPTTPKSAPTTPKEPAPTTPKAPATTPKEPAPTTPKAPATTPKEPAPTTPK 540

Db 440 PTTTPKEPAPTTPKSAPTTPKEPAPTTPKAPATTPKEPAPTTPKAPATTPKEPAPTTPK 499

QY 541 EPAPTTPKAPATTPKEPAPTTPKAPATTPKKTTPPTPEKLAPTTPEKAPATTPPELAP 600

Db 500 EPAPTTPKAPATTPKEPAPTTPKAPATTPKKTTPPTPEKLAPTTPEKAPATTPPELAP 559

QY 601 TTPPEPTTPTEEPAPTTPKAAANTPKAPATTPKEPAPTTPKEPAPTTPKEPAPTTPK 660

Db 560 TTPPEPTTPTEEPAPTTPKAAANTPKAPATTPKEPAPTTPKEPAPTTPKEPAPTTPK 619

QY 661 GTAPTTLKEPAPTTPKAPKELAPTTTKEPTSTSDKAPATTPKGTAPTTPKEPAPTTP 720

Db 620 GTAPTTLKEPAPTTPKAPKELAPTTTKEPTSTSDKAPATTPKGTAPTTPKEPAPTTP 679

QY 721 KEAPATTPKGTAPTTLKEPAPTTPKAPKELAPTTTKEPTSTSDKAPATTPKGTAPTTP 780

Db 680 KEAPATTPKGTAPTTLKEPAPTTPKAPKELAPTTTKEPTSTSDKAPATTPKGTAPTTP 739

QY 781 PKEPAPTTPKAPATTPPTTPPTTSEVSTPTTKEPTTIHKSPDESTPELSAETPKALE 840

Db 740 PKEPAPTTPKAPATTPPTTPPTTSEVSTPTTKEPTTIHKSPDESTPELSAETPKALE 799

QY 841 NSPKFPGVPTTKPAATKPEMTTAKOKTTRDRLRTPPTTAAAPKMTKETATTTETKTE 900

Db 800 NSPKFPGVPTTKPAATKPEMTTAKOKTTRDRLRTPPTTAAAPKMTKETATTTETKTE 859

QY 901 SKITATTTQVSTTTQDTPPKITLTKTTLAPKVTITTKTITTEIMNKPEETAKPKDR 960

Db 860 SKITATTTQVSTTTQDTPPKITLTKTTLAPKVTITTKTITTEIMNKPEETAKPKDR 919

QY 961 ATNKAATTPKOKPTKAPKPTSTTKPKTMPRVKPKTTPTRKMTSTMPELNPTSRIAE 1020

Db 920 ATNKAATTPKOKPTKAPKPTSTTKPKTMPRVKPKTTPTRKMTSTMPELNPTSRIAE 979

QY 1021 AMLQTTTPNQTNSKLVNPKSDAGGAGETPHMLLRPHVMPVETPDMYLPVNP 1080

Db 980 AMLQTTTPNQTNSKLVNPKSDAGGAGETPHMLLRPHVMPVETPDMYLPVNP 1039

QY 1081 QGIINPMLS 1090

Db 1040 QGIINPMLS 1049

RESULT 11

US-10-124-557-142

; Sequence 142, Application US/10124557

; Publication No. US20020137894A1

; GENERAL INFORMATION:

APPLICANT: Turner, Katherine

Clark, Stephen C.

Jacobs, Kenneth

Hewick, Rodney M.

Gesner, Thomas G.

TITLE OF INVENTION: Megakaryocyte Stimulating Factors

NUMBER OF SEQUENCES: 143

CORRESPONDENCE ADDRESS:

ADDRESSEE: Genetics Institute, Inc.

STREET: 87 CambridgePark Drive

CITY: Cambridge

STATE: Massachusetts

COUNTRY: U.S.A.

ZIP: 02140

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patent in Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/10/124,557

FILING DATE: 16-Apr-2002

CLASSIFICATION: <Unknown>

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 07/643,502

FILING DATE: 18-JAN-1991

APPLICATION NUMBER: US 07/546,114

FILING DATE: 29-JUN-1990

APPLICATION NUMBER: US 07/457,196

FILING DATE: 29-DEC-1989

APPLICATION NUMBER: US 07/390,901

FILING DATE: 08-AUG-1989

ATTORNEY/AGENT INFORMATION:

NAME: Cserr, Luann

REGISTRATION NUMBER: 31,822

REFERENCE/DOCKET NUMBER: GI 5190

TELECOMMUNICATION INFORMATION:

TELEPHONE: (617) 876-1170

TELEFAX: (617) 876-5851

INFORMATION FOR SEQ ID NO: 142:

SEQUENCE CHARACTERISTICS:

LENGTH: 1313 amino acids

TYPE: amino acid

TOPOLOGY: linear

MOLSCULE TYPE: protein

SEQUENCE DESCRIPTION: SEQ ID NO: 142:

US-10-124-557-142

Query Match 95.5%; Score 5561.9; DB 13; Length 1313;

Best Local Similarity 96.2%; Pred. No. 1.7e-136;

Matches 1049; Conservative 0; Mismatches 0; Indels 41; Gaps 1;

QY 1 MAWKTLPIYLLLSLVFVIQQVSSQ----- 25

Db 1 MAWKTLPIYLLLSLVFVIQQVSSQ----- 25

QY 61 KRVCYTAELSCGRCFESFERGECDCDAQCKYDKCCPDYBSFCAEHSVSENQESSSS 120

Db 26 -----ELSCGRCFESFERGECDCDAQCKYDKCCPDYBSFCAEHSVSENQESSSS 79

QY 121 SSSSSSTIWKIKSSKNSAANRELOKLVKDNKKNRTKKKPTPKPPVVDAGSLDNGD 180

Db 80 SSSSSSTIWKIKSSKNSAANRELOKLVKDNKKNRTKKKPTPKPPVVDAGSLDNGD 139

QY 181 FKVTTPDSTTCHNKVSTSPKLTAKPINRPSLPNSDTSKETSLSLVNKETTIVETKETT 240

Db 140 FKVTPTDSTTQHNKYSTSPKITTAKPINRPSLSPNSDTSKETSILVNNKETTVEKSTT 199
QY 241 TTNKQSTGKKEKTSIAKTSQSIKTSKADLAPTSKVLAKPTPKAETTTKGPALTTKPEP 300
Db 200 TTNKQSTGKKEKTSIAKTSQSIKTSKADLAPTSKVLAKPTPKAETTTKGPALTTKPEP 259
QY 301 TPTPKPEASTTKEPTPTTIKSAAPTTPKPEAPTITKSAAPTTPKPEAPTITKPEAPTTPK 360
Db 260 TPTPKPEASTTKEPTPTTIKSAAPTTPKPEAPTITKSAAPTTPKPEAPTITKPEAPTTPK 319
QY 361 EPAPTTTKEPAPTITKSAAPTTPKPEAPTITKSAAPTTPKPEAPTITKPEAPTTPKPEAPT 420
Db 320 EPAPTTTKEPAPTITKSAAPTTPKPEAPTITKSAAPTTPKPEAPTITKPEAPTTPKPEAPT 379
QY 421 TKEPAPTTPKPEAPTAPKPEAPTTPKPEAPTTPKPEAPTTPKPEAPTTPKPEAPTTPKSA 480
Db 380 TKEPAPTTPKPEAPTAPKPEAPTTPKPEAPTTPKPEAPTTPKPEAPTTPKPEAPTTPKSA 439
QY 481 PTTTKEPAPTITKSAAPTTPKPEAPTTPKPEAPTTPKPEAPTTPKPEAPTTPKPEAPTTPK 540
Db 440 PTTTKEPAPTITKSAAPTTPKPEAPTTPKPEAPTTPKPEAPTTPKPEAPTTPKPEAPTTPK 499
QY 541 EPAPTTTKEPAPTAPKPEAPTTPKPEAPTTPKPEAPTTPKPEAPTTPKPEAPTTPKPEAPT 600
Db 500 EPAPTTTKEPAPTAPKPEAPTTPKPEAPTTPKPEAPTTPKPEAPTTPKPEAPTTPKPEAPT 559
QY 601 TPEESTPTTPEPAPTTPKAAAPNTPKPEAPTTPKPEAPTTPKPEAPTTPKPEAPTTPK 660
Db 560 TPEESTPTTPEPAPTTPKAAAPNTPKPEAPTTPKPEAPTTPKPEAPTTPKPEAPTTPK 619
QY 661 GTAPTLKEPAPTTPKPKAPKELAPTTTKEPTSTTSKPEAPTTPKGTAPTTPKPEAPTTP 720
Db 620 GTAPTLKEPAPTTPKPKAPKELAPTTTKEPTSTTSKPEAPTTPKGTAPTTPKPEAPTTP 679
QY 721 KPEAPTTPKGTAPTTPKPEAPTTPKPKAPKELAPTTTKEPTSTTSKPEAPTTPKPEAPTTP 780
Db 680 KPEAPTTPKGTAPTTPKPEAPTTPKPKAPKELAPTTTKEPTSTTSKPEAPTTPKPEAPTTP 739
QY 781 KPEAPTTPKPEAPTTPPEPTTSEVSTTSKPEAPTTPKPEAPTTPKPEAPTTPKPEAPTTP 840
Db 740 KPEAPTTPKPEAPTTPPEPTTSEVSTTSKPEAPTTPKPEAPTTPKPEAPTTPKPEAPTTP 799
QY 841 NSPKPEGVTTTTPAATKEMTTAKDKTERDLRTTPTTTAAKPTKETAATTEKTE 900
Db 800 NSPKPEGVTTTTPAATKEMTTAKDKTERDLRTTPTTTAAKPTKETAATTEKTE 859
QY 901 SKITATTTQVSTTTQDTPPKITTLKTTTLAPKVTITKTTTITTEINMKPEETAKPKDR 960
Db 860 SKITATTTQVSTTTQDTPPKITTLKTTTLAPKVTITKTTTITTEINMKPEETAKPKDR 919
QY 961 ATNSKATTPKPKAPKPKPTSTKPKTMPVRKPKTPTPRKMTSTMPBLNPTSRAE 1020
Db 920 ATNSKATTPKPKAPKPKPTSTKPKTMPVRKPKTPTPRKMTSTMPBLNPTSRAE 979
QY 1021 AMLQTTTRENQTPNSKLVNPKSEDAGAGETPHMLLRPHVFMPEVTPDMDYLPVFN 1080
Db 980 AMLQTTTRENQTPNSKLVNPKSEDAGAGETPHMLLRPHVFMPEVTPDMDYLPVFN 1039
QY 1081 QGIINPMLS 1090
Db 1040 QGIINPMLS 1049

RESULT 12
US-10-124-557-52
; Sequence 52, Application US/10124557
; Publication No. US20020137894A1
; GENERAL INFORMATION:
; APPLICANT: Turner, Katherine
; Clark, Stephen C.
; Jacobs, Kenneth
; Hewick, Rodney M.

Gesner, Thomas G.
TITLE OF INVENTION: Megakaryocyte Stimulating Factors
NUMBER OF SEQUENCES: 143
CORRESPONDENCE ADDRESS:
ADDRESSEE: Genetics Institute, Inc.
STREET: 87 Cambridgepark Drive
CITY: Cambridge
STATE: Massachusetts
COUNTRY: U.S.A.
ZIP: 02140
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/10/124,557
FILING DATE: 16-Apr-2002
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/643,502
FILING DATE: 18-JAN-1991
APPLICATION NUMBER: US 07/546,114
FILING DATE: 29-JUN-1990
APPLICATION NUMBER: US 07/457,196
FILING DATE: 29-DEC-1989
APPLICATION NUMBER: US 07/390,901
FILING DATE: 08-AUG-1989
ATTORNEY/AGENT INFORMATION:
NAME: Csery, Luann
REGISTRATION NUMBER: 31,822
REFERENCE/DOCKET NUMBER: GI 5190
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617)876-1170
TELEFAX: (617)876-5851
INFORMATION FOR SEQ ID NO: 52:
SEQUENCE CHARACTERISTICS:
LENGTH: 1363 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
SEQUENCE DESCRIPTION: SEQ ID NO: 52:
US-10-124-557-52

Query Match 95.2%; Score 5546.9; DB 13; Length 1363;
Best Local Similarity 92.0%; Pred. No. 4.4e-136;
Matches 1049; Conservative 0; Mismatches 0; Indels 91; Gaps 2;
QY 1 MAWKTLPIYLLLLSVFVIQVSSQDLSSCAGRCGEGYSRDATAICNDYNCQHYMECCPDF 60
Db 1 MAWKTLPIYLLLLSVFVIQVSSQ----- 25
QY 61 KRVCATLSCKGRCFESFERGREGDCDCAOCKKYDKCCPDYESFCA----- 105
Db 26 -----ELSCCKGRCFESFERGREGDCDCAOCKKYDKCCPDYESFCAEVHNFTSPSSKAP 79
QY 106 -----BEHSVSENOESSSSSSSSSSSSSSSIW 130
Db 80 PPSGASOTIKSTTKESPKPNKKTKVIESEITEHSVSENOESSSSSSSSSSSSSIW 139
QY 131 KIKSKNSAANRELQKLVKDKNKNRTKKKPTPKPPVDEAGSLDNGDFKVTTPDTST 190
Db 140 KIKSKNSAANRELQKLVKDKNKNRTKKKPTPKPPVDEAGSLDNGDFKVTTPDTST 199
QY 191 TQHNKVTSPKITTAKPINRPSLPNSDTSKETSILVNNKETTVEKETTNNKQSTSDG 250
Db 200 TQHNKVTSPKITTAKPINRPSLPNSDTSKETSILVNNKETTVEKETTNNKQSTSDG 259
QY 251 KEKTTSAKETSQSIKTSIAKTSKADLAPTSKVLAKPTPKAETTTKGPALTTKPEPTTPPKPEAS 310
Db 260 KEKTTSAKETSQSIKTSIAKTSKADLAPTSKVLAKPTPKAETTTKGPALTTKPEPTTPPKPEAS 319
QY 311 TTPKEPTTIKSAPTTPKPEAPTTPKSAPTTPKPEAPTTPKPEAPTTPKPEAPTTPK 370

498 TAFTTPEKLTFTTPEKAPATTPEKAPATTPEELAPTTPEEPTFTTPEEAPATTPKAAAPN 557
626 TPKEPAPATTPEKAPATTPEKAPATTPEKAPATTPEKAPATTPEKAPATTPEKAPATTPEKAP 685
558 TPKEPAPATTPEKAPATTPEKAPATTPEKAPATTPEKAPATTPEKAPATTPEKAPATTPEKAP 617
686 TTKKEPTSTSDKAPATTPEKAPATTPEKAPATTPEKAPATTPEKAPATTPEKAPATTPEKAP 745
618 TTKKEPTSTSDKAPATTPEKAPATTPEKAPATTPEKAPATTPEKAPATTPEKAPATTPEKAP 677
746 KPAPKELAPTTKGPSTSTSDKAPATTPEKAPATTPEKAPATTPEKAPATTPEKAPATTPEKAP 805
678 KPAPKELAPTTKGPSTSTSDKAPATTPEKAPATTPEKAPATTPEKAPATTPEKAPATTPEKAP 737
806 EVSTPTTKEPTTHKSPDESPELSAEPTEKALENSPKPEPGVPTTKPAATKPSWTTTA 865
738 EVSTPTTKEPTTHKSPDESPELSAEPTEKALENSPKPEPGVPTTKPAATKPSWTTTA 797
866 KOKTTERDLRTPEPTTAAAPKMTKETAATTTKTESKITATTTQVTSITTTQDTTFFKITT 925
798 KOKTTERDLRTPEPTTAAAPKMTKETAATTTKTESKITATTTQVTSITTTQDTTFFKITT 857
926 LKTTTLAPKVTTKTITTTTEIMNKPEETAKPKDRATNSKATTPKQKTKAPKPTSTK 985
858 LKTTTLAPKVTTKTITTTTEIMNKPEETAKPKDRATNSKATTPKQKTKAPKPTSTK 917
986 KPKTMRVRKPKTTTPPRKWTSTMPPELNPSTRIAEAMLQTTTRPNQTNPKSLVEVNPXSE 1045
918 KPKTMRVRKPKTTTPPRKWTSTMPPELNPSTRIAEAMLQTTTRPNQTNPKSLVEVNPXSE 977
1046 DAGGAEGETPHMLLRPHVMPVEPTDMDYLPRVNPQGIINPMLS 1090
978 DAGGAEGETPHMLLRPHVMPVEPTDMDYLPRVNPQGIINPMLS 1022

RESULT 14.

US-10-124-557-46
; Sequence 46, Application US/10124557
; Publication No. US20020137894A1
; GENERAL INFORMATION:
; APPLICANT: Turner, Katherine
; Clark, Stephen C.
; Jacobs, Kenneth
; Hewick, Rodney M.
; Gesner, Thomas G.
; TITLE OF INVENTION: Megakaryocyte Stimulating Factors
; NUMBER OF SEQUENCES: 143
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genetics Institute, Inc.
; STREET: 87 CambridgePark Drive
; CITY: Cambridge
; STATE: Massachusetts
; COUNTRY: U.S.A.
; ZIP: 02140
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION NUMBER: US/10/124,557
; APPLICATION DATA:
; FILING DATE: 16-Apr-2002
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/643,502
; FILING DATE: 18-JAN-1991
; APPLICATION NUMBER: US 07/546,114
; FILING DATE: 29-JUN-1990
; APPLICATION NUMBER: US 07/457,196
; FILING DATE: 29-DEC-1989
; APPLICATION NUMBER: US 07/390,901
; FILING DATE: 08-AUG-1989

ATTORNEY/AGENT INFORMATION:
; NAME: Cseriz, Luanh
; REGISTRATION NUMBER: 31,822
; REFERENCE/DOCKET NUMBER: GI 5190
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617) 876-1170
; TELEFAX: (617) 876-5851
; INFORMATION FOR SEQ ID NO: 46:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1320 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; SEQUENCE DESCRIPTION: SEQ ID NO: 46:
US-10-124-557-46
Query Match 92.3%; Score 5379.2; DB 13; Length 1320;
Best Local Similarity 92.5%; Pred. No. 9.5e-132;
Matches 1015; Conservative 8; Mismatches 26; Indels 48; Gaps 2;
QY 1 MAWKTLPIYLLLLSVFVIOQVSSQDLSSCAGRCGEGYSRDATCNDYCNQHYMECCPDF 60
DB 1 MAWKTLPIYLLLLSVFVIOQVSSQ----- 25
QY 61 KRVCTAELSCKGRCFESFERGECDDAQCCKYDKCCPDYESFCAEHSVSENQESSSS 120
DB 26 -----ELSCKGRCFESFERGECDDAQCCKYDKCCPDYESFCAEHSVSENQESSKAP 79
QY 121 SSSSSSTIWKIKSSKNSAANRELQKL-----KVONKKNKRTKKKPTKPPVVDAG 173
DB 80 PPGASGTIIKSTTKRSPKPPNKKTKRVIESEITTEVKNKKNKRTKKKPTKPPVVDAG 139
QY 174 SGLDNGDFKVTTPDTSTTHNKNVSTSPKITTAKPINPRPSLPNPSDTSKETSITVNKETT 233
DB 140 SGLDNGDFKVTTPDTSTTHNKNVSTSPKITTAKPINPRPSLPNPSDTSKETSITVNKETT 199
QY 234 VEYKETTITNKQSTDGKEKITSAKETQSIEKTSKADLAPTSKVLAKPTKAEITTKGPA 293
DB 200 VETKEITTTNKQSTDGKEKITSAKETQSIEKTSKADLAPTSKVLAKPTKAEITTKGPA 259
QY 294 LITPKKEPTTPPKPEPASTTTPKEPTTIKSAPTTPKEPAPATTTKSAPTTPKEPAPATTKE 353
DB 260 LITPKKEPTTPPKPEPASTTTPKEPTTIKSAPTTPKEPAPATTTKSAPTTPKEPAPATTKE 319
QY 354 PAPTTPKEPAPATTTKSAPTTPKEPAPATTTKSAPTTPKEPAPATTTKSAPTTPKEPAPATT 413
DB 320 PAPTTPKEPAPATTTKSAPTTPKEPAPATTTKSAPTTPKEPAPATTTKSAPTTPKEPAPATT 379
QY 414 PKPEPAPATTKEPAPATTKEPAPATTKEPAPATTKEPAPATTKEPAPATTKEPAPATTKEP 473
DB 380 PKPEPAPATTKEPAPATTKEPAPATTKEPAPATTKEPAPATTKEPAPATTKEPAPATTKEP 439
QY 474 PTTTKSAPTTPTKSAPTTPKEPAPATTTKSAPTTPKEPAPATTTKSAPTTPKEPAPATTKE 533
DB 440 PTTTKSAPTTPTKSAPTTPKEPAPATTTKSAPTTPKEPAPATTTKSAPTTPKEPAPATTKE 499
QY 534 PAPTTPKEPAPATTTKSAPTTPKEPAPATTTKSAPTTPKEPAPATTTKSAPTTPKEPAPATT 593
DB 500 PAPTTPKEPAPATTTKSAPTTPKEPAPATTTKSAPTTPKEPAPATTTKSAPTTPKEPAPATT 559
QY 594 TPBELAPTTPEEPTTPPEEPAPATTPKAAAPNTPKPEPAPATTPKPEPAPATTPKPEPAPATT 653
DB 560 TPBELAPTTPEEPTTPPEEPAPATTPKAAAPNTPKPEPAPATTPKPEPAPATTPKPEPAPATT 619
QY 654 TAPTTPKGTAPTTLKBPAPATTPKKAPKELAPTTTKEPTSTTSKDPAPATTPKGTAPTTPK 713
DB 620 TAPTTPKGTAPTTLKBPAPATTPKKAPKELAPTTTKEPTSTTSKDPAPATTPKGTAPTTPK 679
QY 714 EPAPTTPKEPAPATTTPKGTAPTTLKBPAPATTPKKAPKELAPTTTKEPTSTTSKDPAPATT 773
DB 680 EPAPTTPKEPAPATTTPKGTAPTTLKBPAPATTPKKAPKELAPTTTKEPTSTTSKDPAPATT 739
QY 774 KETAPTTPEPAPATTPKKAPKELAPTTTKEPTSTTSKDPAPATTTKEPTSTTSKDPAPATT 833

DB 740 KETAPTPKEAPTPKPAAPTPETPTPTTSEVSTPTTKEPTTIHKSPDESPELSAE 799
QY 834 PTPKALENSPKPEGVPPTTKTAAATKPEMTTAKOKTTERDLRTTPEITTAAPKMTKETAT 893
DB 800 PTPKALENSPKPEGVPPTTKTAAATKPEMTTAKOKTTERDLRTTPEITTAAPKMTKETAT 859
QY 894 TTEKTTESKITATTQVSTTTQDTPPKIITTLKTIILAPKVTTTKIITTEIMNKPEE 953
DB 860 TTEKTTESKITATTQVSTTTQDTPPKIITTLKTIILAPKVTTTKIITTEIMNKPEE 919
QY 954 TAKPKDRATNSKATPKPKPTKAPKPTSTKPKTWPVRVKPTTTPRKMWTSTMPELN 1013
DB 920 TAKPKDRATNSKATPKPKPTKAPKPTSTKPKTWPVRVKPTTTPRKMWTSTMPELN 979
QY 1014 PPSRIAEAMLQTTTPRNPNSKLVNPNKSEDAGAEGETPHMLLRPHVFMPEVTPDMD 1073
DB 980 PPSRIAEAMLQTTTPRNPNSKLVNPNKSEDAGAEGETPHMLLRPHVFMPEVTPDMD 1039
QY 1074 YLPRVNOGIIINPMLS 1090
DB 1040 YLPRVNOGIIINPMLS 1056

RESULT 15
US-10-124-557-60
; Sequence 60, Application US/10124557
; Publication No. US20020137894A1
; GENERAL INFORMATION:
; APPLICANT: Turner, Katherine
; Jacobs, Kenneth
; Clark, Stephen C.
; Hewick, Rodney M.
; Gesner, Thomas G.
; TITLE OF INVENTION: Megakaryocyte Stimulating Factors
; NUMBER OF SEQUENCES: 143
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genetics Institute, Inc.
; STREET: 87 CambridgePark Drive
; CITY: Cambridge
; STATE: Massachusetts
; COUNTRY: U.S.A.
; ZIP: 02140
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA: US/10/124,557
; FILING DATE: 16-Apr-2002
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/643,502
; FILING DATE: 18-JAN-1991
; APPLICATION NUMBER: US 07/546,114
; FILING DATE: 29-JUN-1990
; APPLICATION NUMBER: US 07/457,196
; FILING DATE: 29-DEC-1989
; APPLICATION NUMBER: US 07/390,901
; FILING DATE: 08-AUG-1989
; ATTORNEY/AGENT INFORMATION:
; NAME: Cseri, Luann
; REGISTRATION NUMBER: 31,822
; REFERENCE/DOCKET NUMBER: GI 5190
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617)876-1170
; TELEFAX: (617)876-5851
; INFORMATION FOR SEQ ID NO: 60:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1320 amino acids
; TYPE: amino acid
; TOPOLOGY: linear

MOLECULE TYPE: protein
SEQUENCE DESCRIPTION: SEQ ID NO: 60:
US-10-124-557-60
Query Match 92.3%; Score 5379.2; DB 13; Length 1320;
Best Local Similarity 92.5%; Pred. No. 9.5e-132;
Matches 1015; Conservative 8; Mismatches 26; Indels 48; Gaps 2;
QY 1 MANKTLPIYLLALLLVFVIQQVSSQDLSSCAGRCGEGYSRDATCNDYNCQHMYMCCPDF 60
DB 1 MANKTLPIYLLALLLVFVIQQVSSQ----- 25
QY 61 KRVTAEALCKGRCEPESFERGRECDCAQCKKYDKCCPDYSEFCABEHSVSNQSSSS 120
DB 26 -----ELSKGRCEPESFERGRECDCAQCKKYDKCCPDYSEFCABEHSVSNQSSSSKKAP 79
QY 121 SSSSSSTTWKTSKNSAANRELQKL-----KVDNKNRTPKKKPTKPPVVDGAG 173
DB 80 PPSGASQTIKSTTKSPKPFNKKTKKVISEEITEVDNKNRTPKKKPTKPPVVDGAG 139
QY 174 SGLDNGDFKVTTPDSTTQHNKVSTSPKITTAKPINPRPSLPNSDTSKETSITVKNKETT 233
DB 140 SGLDNGDFKVTTPDSTTQHNKVSTSPKITTAKPINPRPSLPNSDTSKETSITVKNKETT 199
QY 234 VETKETTNTNKTSDGKEKTSKETSQSIETSAKOLAPTSKVLAKPTPKAETTTKGA 293
DB 200 VETKETTNTNKTSDGKEKTSKETSQSIETSAKOLAPTSKVLAKPTPKAETTTKGA 259
QY 294 LITPKETPTTPKEPASTTPKEPTPTTIKASATTPKEPAPTTKSAPTTPKEPAPTTTKE 353
DB 260 LITPKETPTTPKEPASTTPKEPTPTTIKASATTPKEPAPTTKSAPTTPKEPAPTTTKE 319
QY 354 PAPTTPKEPAPTTTPKEPAPTTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPTPTT 413
DB 320 PAPTTPKEPAPTTTPKEPAPTTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPTPTT 379
QY 414 PKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAP 473
DB 380 PKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAP 439
QY 474 PTTTKSAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKE 533
DB 440 PTTTKSAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKE 499
QY 534 PAPTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKE 593
DB 500 PAPTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKE 559
QY 594 TPBELAPTTPEEPTPTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTT 653
DB 560 TPBELAPTTPEEPTPTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTT 619
QY 654 TAPTTPKGATPTTLKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKE 713
DB 620 TAPTTPKGATPTTLKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKE 679
QY 714 EPAPTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTP 773
DB 680 EPAPTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTP 739
QY 774 KETAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTT 833
DB 740 KETAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTT 799
QY 834 PTPKALENSPKPEGVPPTTKTAAATKPEMTTAKOKTTERDLRTTPEITTAAPKMTKETAT 893
DB 800 PTPKALENSPKPEGVPPTTKTAAATKPEMTTAKOKTTERDLRTTPEITTAAPKMTKETAT 859
QY 894 TTEKTTESKITATTQVSTTTQDTPPKIITTLKTIILAPKVTTTKIITTEIMNKPEE 953
DB 860 TTEKTTESKITATTQVSTTTQDTPPKIITTLKTIILAPKVTTTKIITTEIMNKPEE 919
QY 954 TAKPKDRATNSKATPKPKPTKAPKPTSTKPKTWPVRVKPTTTPRKMWTSTMPELN 1013

| | | | |
|----|------|--|------|
| Db | 920 | TAKPKDRATNSKATTPKQKPTKAPKPTSTKKPTMPRVKPKTTPRKMSTMPELN | 979 |
| Qy | 1014 | PTSRIAEMLOTTTRPNOTPNSKLYEVNPKSEDAAGAGETPHMLLRPHVFMPEVTPDMD | 1073 |
| Db | 980 | PTSRIAEMLOTTTRPNOTPNSKLYEVNPKSEDAAGAGETPHMLLRPHVFMPEVTPDMD | 1039 |
| Qy | 1074 | YLPRVFNQGIINPMLS | 1090 |
| Db | 1040 | YLPRVFNQGIINPMLS | 1056 |

Search completed: October 13, 2004, 11:52:45
Job time : 117.895 secs

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: October 13, 2004, 11:23:49 ; Search time 119.845 Seconds
(without alignments)
5233.063 Million cell updates/sec

Title: SEQ1-B
Perfect score: 5826
Sequence: 1 MAWKTLPIYLLLLSVFVIQ.....DMDYLPVPNOGIIINPMLS 1090

Scoring table: BLOSUM62
Gapop 10.0., Gapext 0.1

Searched: 1825181 seqs, 575374646 residues

Total number of hits satisfying chosen parameters: 1825181

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : UniProt 02: *
1: uniprot_sprot: *
2: uniprot_trembl: *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | ID | Description |
|------------|--------|-------------|--------|--------------|---------------------|
| 1 | 5811 | 99.7 | 1404 | 2 Q92954 | Q92954 homo sapien |
| 2 | 5788 | 99.3 | 1404 | 2 Q9BX49 | Q9BX49 homo sapien |
| 3 | 3128.9 | 53.7 | 933 | 2 Q62MZ5 | Q62MZ5 homo sapien |
| 4 | 3128.9 | 53.7 | 933 | 2 BAD18580 | BAD18580 h cdna fl |
| 5 | 2700.5 | 46.4 | 1054 | 2 Q9JN99 | Q9JN99 mus musculus |
| 6 | 1394.5 | 23.9 | 5179 | 1 MUC2_HUMAN | MUC2_HUMAN |
| 7 | 1321.9 | 22.7 | 1225 | 2 Q9VR49 | Q9VR49 drosophila |
| 8 | 1314.1 | 22.6 | 1761 | 2 Q7KTF6 | Q7KTF6 drosophila |
| 9 | 1314.1 | 22.6 | 1761 | 2 AAS64673 | AAS64673 drosophila |
| 10 | 1243.6 | 21.3 | 3150 | 2 Q7PMD5 | Q7PMD5 anopheles g |
| 11 | 1187.5 | 20.4 | 3409 | 2 Q6SSE6 | Q6SSE6 chlamydomon |
| 12 | 1187.5 | 20.4 | 3409 | 2 AAS07044 | AAS07044 chlamydom |
| 13 | 1127.2 | 19.3 | 1664 | 1 SLPI_CLOTM | SLPI_CLOTM |
| 14 | 1113.7 | 19.1 | 1349 | 2 Q8WQ04 | Q8WQ04 homo sapien |
| 15 | 1103.7 | 18.9 | 1795 | 2 Q76S94 | Q76S94 drosophila |
| 16 | 1098.1 | 18.8 | 3889 | 2 Q6SSE8 | Q6SSE8 chlamydomon |
| 17 | 1098.1 | 18.8 | 3889 | 2 AAS07042 | AAS07042 chlamydom |
| 18 | 1090.2 | 18.7 | 3432 | 2 Q8IR51 | Q8IR51 drosophila |
| 19 | 1090.2 | 18.7 | 3432 | 2 Q8IR52 | Q8IR52 drosophila |
| 20 | 1059.8 | 18.2 | 1079 | 2 Q9N4S7 | Q9N4S7 caenorhabdi |
| 21 | 1035.3 | 17.8 | 9234 | 2 Q7KTF5 | Q7KTF5 drosophila |
| 22 | 1035.3 | 17.8 | 9234 | 2 AAN10531 | AAN10531 drosophila |
| 23 | 1031 | 17.7 | 2284 | 2 Q9VPG1 | Q9VPG1 drosophila |
| 24 | 1028.7 | 17.7 | 1607 | 2 Q6HQ05 | Q6HQ05 phytophthor |
| 25 | 1028.7 | 17.7 | 1607 | 2 AAP74661 | AAP74661 phytophth |
| 26 | 1027 | 17.6 | 5703 | 1 MUSE_HUMAN | MUSE_HUMAN |
| 27 | 1022.7 | 17.6 | 972 | 2 Q7QKK7 | Q7QKK7 anopheles g |
| 28 | 1019.8 | 17.5 | 23015 | 2 Q8IQ18 | Q8IQ18 drosophila |
| 29 | 1019.8 | 17.5 | 23015 | 2 AAN10358 | AAN10358 drosophila |
| 30 | 1004.7 | 17.2 | 1489 | 2 Q96449 | Q96449 phytophthor |
| 31 | 1002.8 | 17.2 | 1274 | 2 Q20007 | Q20007 caenorhabdi |

RESULT 1

Q92954 PRELIMINARY; PRT; 1404 AA.
AC Q92954
DT 01-FEB-1997 (TrEMBLrel. 02, Created)
DT 01-FEB-1997 (TrEMBLrel. 02, Last sequence update)
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE Megakaryocyte stimulating factor.
OS Homo sapiens (Human)
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Turner K.J., Fitz L.J., Temple P., Jacobs K., Larson D., Leary A.C.,
RA Kelleher K., Giannotti J., Calvetti J., Fitzgerald M., Kriz M.J.,
RA Ferenz C., Grobholz J., Fraser H., Bean K., Norton C.R., Gesner T.,
RA Bhatia S., Kriz R., Hewick R., Fricker R., Clark S.C.;
RT "Purification, Biochemical Characterization, and Cloning of a Novel
RT Megakaryocyte Stimulating Factor that has Megakaryocyte Colony
RT Stimulating Activity";
RL Blood 78:279-279(1991).
RN [2]
RP SEQUENCE FROM N.A.
RA Merberg D.M., Fitz L.J., Temple P., Giannotti J., Murtha P.,
RA Fitzgerald M., Scattreto J., Kelleher K., Freissner K., Kriz R.,
RA Jacobs K., Turner K.;
RT "A Comparison of Vitronectin and Megakaryocyte Stimulating Factor";
RL Mosher D.F. (eds.);
RL BIOLOGY OF VITRONECTINS AND THEIR RECEPTORS., pp.45-52, Elsevier
RL Science Publishers B.V. (1993).
RN [3]
RP SEQUENCE FROM N.A.
RA Turner K.J., Fitz L.J., Temple P., Jacobs K., Larson D., Leary A.C.,
RA Kelleher K., Giannotti J., Calvetti J., Fitzgerald M., Kriz M.J.,
RA Ferenz C., Grobholz J., Fraser H., Bean K., Norton C.R., Gesner T.,
RA Bhatia S., Kriz R., Hewick R., Fricker R., Clark S.C.;
RT Submitted (SEP-1996) to the EMBL/GenBank/DBJ databases.
RL EMBL; U70136; AAB09089.1; ;
RL HSP; P04004; I0C0.
RL Genew; HGNC:9364; PRG4.
GO; GO:0008283; P:cell proliferation; TAS.
DR InterPro; IPR000585; Hemopexin.
DR InterPro; IPR001212; Somatomedin_B.
DR Pfam; PF00045; Hemopexin; 2.
DR Pfam; PF01033; Somatomedin B; 2.
DR PRINTS; PR00022; SOMATOMEDINB.
DR SMART; SM00120; HX; 2.
DR SMART; SM00201; SO; 2.
DR PROSITE; PS00024; HEMOPEXIN; UNKNOWN_1.
DR PROSITE; PS00524; SOMATOMEDIN_B; 2.
SQ SEQUENCE 1404 AA; 151090 MW; AABD7AD19B35F4F6 CRC64;

Q9VEL9 drosophila
Q8WZ42 homo sapien
Q10465 homo sapien
Q8ISF5 caenorhabdi
Q8ISF5 caenorhabdi
Q8ISF6 caenorhabdi
Q8ISF7 caenorhabdi
P70670 mus musculus
Q9VVG2 drosophila
Q757N5 ashbya goss
AAS52662 ashbya go
Q81486 plasmodium
Q76718 sus scrofa
BAD08434 sus scrofa
Q7Z884 candida alb

ALIGNMENTS

Query Match 99.7%; Score 5811; DB 2; Length 1404;
Best Local Similarity 95.6%; Pred. No. 9,6e-106;
Matches 1090; Conservative 0; Mismatches 0; Indels 50; Gaps 1;

QY 1 MAWKTLPIVLLLLSVFVIQQVSSQDLSSCAGRCGEGYSDATCNCYNQHYMECCPDF 60
DB 1 MAWKTLPIVLLLLSVFVIQQVSSQDLSSCAGRCGEGYSDATCNCYNQHYMECCPDF 60

QY 61 KRVTAEKSCGRCFESFERGECDCDAQCKYDKCCPDYESFCA----- 105
DB 61 KRVTAEKSCGRCFESFERGECDCDAQCKYDKCCPDYESFCAEVHNPSPSSKKAP 120

QY 106 -----BEHSVSNQSSSSSSSSSSSSSSSIW 130
DB 121 PPSGASQTIKSTTKRSPKPNKKTKKVIKESBEITEHSVSNQSSSSSSSSSSSIW 180

QY 131 KIKSSKNSAANRELOKLVKDNKNRTKKKTPKPPVVDVDEAGSLDNGDFKVTTPDST 190
DB 181 KIKSSKNSAANRELOKLVKDNKNRTKKKTPKPPVVDVDEAGSLDNGDFKVTTPDST 240

QY 191 TOHNVSTSPKITTAKPINRPSLPNSDTSKETSILTNNKETTVEKETTINKQTSIDG 250
DB 241 TOHNVSTSPKITTAKPINRPSLPNSDTSKETSILTNNKETTVEKETTINKQTSIDG 300

QY 251 KEKTTSAKETQSIKTSKADLAPTSKVLAKPTPKAETTTKGPALTTPKEPTTPKEPAS 310
DB 301 KEKTTSAKETQSIKTSKADLAPTSKVLAKPTPKAETTTKGPALTTPKEPTTPKEPAS 360

QY 311 TTPKEPTPTTIKSAPTTPKEPATTPKAPATTPKBPATTPKBPATTPKBPATTPK 370
DB 361 TTPKEPTPTTIKSAPTTPKEPATTPKAPATTPKBPATTPKBPATTPKBPATTPK 420

QY 371 APTTTKSAPTTPKEPATTPKBPATTPKBPATTPKBPATTPKBPATTPKBPATTPK 430
DB 421 APTTTKSAPTTPKEPATTPKBPATTPKBPATTPKBPATTPKBPATTPKBPATTPK 480

QY 431 EPAPTAPKPPAPTTPKEPATTPKBPATTPKBPATTPKBPATTPKBPATTPKBPATTPK 490
DB 481 EPAPTAPKPPAPTTPKEPATTPKBPATTPKBPATTPKBPATTPKBPATTPKBPATTPK 540

QY 491 TTKSAPTTPKESPTTPKEPATTPKBPATTPKBPATTPKBPATTPKBPATTPKBPATTPK 550
DB 541 TTKSAPTTPKESPTTPKEPATTPKBPATTPKBPATTPKBPATTPKBPATTPKBPATTPK 600

QY 551 APTAPKEPATTPKETAPTTPKPLTPTPKLAPTTPKEPATTPKBPATTPKBPATTPK 610
DB 601 APTAPKEPATTPKETAPTTPKPLTPTPKLAPTTPKEPATTPKBPATTPKBPATTPK 660

QY 611 PEEAPTTPKAAAPTTPKEPATTPKBPATTPKBPATTPKBPATTPKBPATTPKBPATTPK 670
DB 661 PEEAPTTPKAAAPTTPKEPATTPKBPATTPKBPATTPKBPATTPKBPATTPKBPATTPK 720

QY 671 APTTPKAPKELAPTTTKEPTSTSDKAPTTPKGTAPTTPKBPATTPKBPATTPKBPATTPK 730
DB 721 APTTPKAPKELAPTTTKEPTSTSDKAPTTPKGTAPTTPKBPATTPKBPATTPKBPATTPK 780

QY 731 TAPTTLKBPATTPKPPAPKELAPTTTGGTSTSDKAPTTPKETAPTTPKBPATTPKBPATTPK 790
DB 781 TAPTTLKBPATTPKPPAPKELAPTTTGGTSTSDKAPTTPKETAPTTPKBPATTPKBPATTPK 840

QY 791 KPAPTTPPTTSEVSTPTTKETPTTIHKSPDESPELSAETPKALENSPKBPVPT 850
DB 841 KPAPTTPPTTSEVSTPTTKETPTTIHKSPDESPELSAETPKALENSPKBPVPT 900

QY 851 TKTPAATKPEMTTAKDXTTERDLRTTPETTTAAPKMTKETATTTETKTTESKTTATTTQV 910
DB 901 TKTPAATKPEMTTAKDXTTERDLRTTPETTTAAPKMTKETATTTETKTTESKTTATTTQV 960

QY 911 TSTTQDTPPKITLTKTTTLAPKVTTTKTITTEIMNKPEETAKPKDRAATNSKATPK 970
DB 961 TSTTQDTPPKITLTKTTTLAPKVTTTKTITTEIMNKPEETAKPKDRAATNSKATPK 1020

QY 971 POKPTAPKPKSTSKPKMTMPVRKPKTTPTPRKMSTMPDELNPTSRIAEAMLTQTTTRN 1030

DB 1021 POKPTAPKPKSTSKPKMTMPVRKPKTTPTPRKMSTMPDELNPTSRIAEAMLTQTTTRN 1080

QY 1031 QTPNSKLVEVNPXSSEDAGGAEGTTPHMLLRPHVPMPEVTPDMDYLRVNPQGGIIINPMLS 1090

DB 1081 QTPNSKLVEVNPXSSEDAGGAEGTTPHMLLRPHVPMPEVTPDMDYLRVNPQGGIIINPMLS 1140

RESULT 2

Q9BX49 PRELIMINARY; PRT; 1404 AA.

ID Q9BX49

AC Q9BX49

DT 01-JUN-2001 (T-EMBLrel. 17, Created)

DT 01-JUN-2001 (T-EMBLrel. 17, Last sequence update)

DT 01-JUN-2004 (T-EMBLrel. 26, Last annotation update)

DE BG174L6.2 (MSF: megakaryocyte stimulating factor).

GN Name=BG174L6.2;

OS Homo sapiens (Human)

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

OC NCBI_TaxID=9606;

RN [1]

RP SEQUENCE FROM N.A.

RA May P.,

RL Submitted (JUL-2000) to the EMBL/GenBank/DBJ databases.

DR EMBL; AL133553; CAC36090.1; -;

DR HSP; P04004; 10C0.

DR InterPro; IPR000585; Hemopexin.

DR InterPro; IPR001212; Somatomedin_B.

DR Pfam; PF00045; Hemopexin; 2.

DR Pfam; PF01033; Somatomedin_B; 2.

DR PRINTS; PR00022; SOMATOMEDINB.

DR SMART; SM00120; HK; 2.

DR SMART; SM00201; SO; 2.

DR PROSITE; PS00024; HEMOPEXIN; UNKNOWN_1.

DR PROSITE; PS00524; SOMATOMEDIN_B; 2.

SQ SEQUENCE 1404 AA; 151076 MW; 782A11746B3FDEB5 CRC64;

Query Match 99.3%; Score 5788; DB 2; Length 1404;
Best Local Similarity 95.4%; Pred. No. 2,7e-105;
Matches 1087; Conservative 0; Mismatches 3; Indels 50; Gaps 1;

QY 1 MAWKTLPIVLLLLSVFVIQQVSSQDLSSCAGRCGEGYSDATCNCYNQHYMECCPDF 60
DB 1 MAWKTLPIVLLLLSVFVIQQVSSQDLSSCAGRCGEGYSDATCNCYNQHYMECCPDF 60

QY 61 KRVTAEKSCGRCFESFERGECDCDAQCKYDKCCPDYESFCA----- 105
DB 61 KRVTAEKSCGRCFESFERGECDCDAQCKYDKCCPDYESFCAEVHNPSPSSKKAP 120

QY 106 -----BEHSVSNQSSSSSSSSSSSSSSSIW 130
DB 121 PPSGASQTIKSTTKRSPKPNKKTKKVIKESBEITEHSVSNQSSSSSSSSSSSIW 180

QY 131 KIKSSKNSAANRELOKLVKDNKNRTKKKTPKPPVVDVDEAGSLDNGDFKVTTPDST 190
DB 181 KIKSSKNSAANRELOKLVKDNKNRTKKKTPKPPVVDVDEAGSLDNGDFKVTTPDST 240

QY 191 TOHNVSTSPKITTAKPINRPSLPNSDTSKETSILTNNKETTVEKETTINKQTSIDG 250
DB 241 TOHNVSTSPKITTAKPINRPSLPNSDTSKETSILTNNKETTVEKETTINKQTSIDG 300

QY 251 KEKTTSAKETQSIKTSKADLAPTSKVLAKPTPKAETTTKGPALTTPKEPTTPKEPAS 310
DB 301 KEKTTSAKETQSIKTSKADLAPTSKVLAKPTPKAETTTKGPALTTPKEPTTPKEPAS 360

QY 311 TTPKEPTPTTIKSAPTTPKEPATTPKAPATTPKBPATTPKBPATTPKBPATTPK 370
DB 361 TTPKEPTPTTIKSAPTTPKEPATTPKAPATTPKBPATTPKBPATTPKBPATTPK 420

QY 371 APTTTKSAPTTPKEPATTPKBPATTPKBPATTPKBPATTPKBPATTPKBPATTPK 430
DB 421 APTTTKSAPTTPKEPATTPKBPATTPKBPATTPKBPATTPKBPATTPKBPATTPK 480

| | | | |
|----|------|--|------|
| QY | 431 | EPAPAPKAPKAPATTTKEBPATTPKBPATTTKEBPATTPKBPATTTKSAPTITKEBPAPT | 490 |
| DB | 481 | EPAPAPKAPKAPATTTKEBPATTPKBPATTTKEBPATTPKBPATTTKSAPTITKEBPAPT | 540 |
| QY | 491 | TTKSAPTTPKBPSPITTTKEBPATTPKBPATTPKBPATTPKBPATTPKBPATTTTKKP | 550 |
| DB | 541 | TTKSAPTTPKBPSPITTTKEBPATTPKBPATTPKBPATTPKBPATTPKBPATTTTKKP | 600 |
| QY | 551 | APTAPKEBPATTPKBPATTPKBPATTPKBPATTPKBPATTPKBPATTPKBPATTTT | 610 |
| DB | 601 | APTTPKEBPATTPKBPATTPKBPATTPKBPATTPKBPATTPKBPATTPKBPATTTT | 660 |
| QY | 611 | PREPAPTPKAAAPNTKBPATTPKBPATTPKBPATTPKBPATTPKBPATTPKBPATTT | 670 |
| DB | 661 | PREPAPTPKAAAPNTKBPATTPKBPATTPKBPATTPKBPATTPKBPATTPKBPATTT | 720 |
| QY | 671 | APTTPKAPKAPKAPATTTKEBPATTPKBPATTPKBPATTPKBPATTPKBPATTTT | 730 |
| DB | 721 | APTTPKAPKAPKAPATTTKEBPATTPKBPATTPKBPATTPKBPATTPKBPATTTT | 780 |
| QY | 731 | TAPTITKEBPATTPKBPATTPKBPATTPKBPATTPKBPATTPKBPATTPKBPATTPK | 790 |
| DB | 781 | TAPTITKEBPATTPKBPATTPKBPATTPKBPATTPKBPATTPKBPATTPKBPATTPK | 840 |
| QY | 791 | KPAPTTPETPTTSEVSTPTTKEPTTIHKSPDESTPELSAEPPTKALENSPKBPVPT | 850 |
| DB | 841 | KPAPTTPETPTTSEVSTPTTKEPTTIHKSPDESTPELSAEPPTKALENSPKBPVPT | 900 |
| QY | 851 | TKTPAATKPEMTTAKDKTTERDLRTPTTAAAPKMTKETATTTKEKTESKITATTQV | 910 |
| DB | 901 | TKTPAATKPEMTTAKDKTTERDLRTPTTAAAPKMTKETATTTKEKTESKITATTQV | 960 |
| QY | 911 | TSVTTQDTPFKITLTKTTLAPKVTITTKTITTTIMNKPEETAKPKDRATNSKATTPK | 970 |
| DB | 961 | TSVTTQDTPFKITLTKTTLAPKVTITTKTITTTIMNKPEETAKPKDRATNSKATTPK | 1020 |
| QY | 971 | POKTPKAPKPTSTKPKTMPVRKPTTPTRKMTSTWPELNPTSRIAEAMLOTTTPN | 1030 |
| DB | 1021 | POKTPKAPKPTSTKPKTMPVRKPTTPTRKMTSTWPELNPTSRIAEAMLOTTTPN | 1080 |
| QY | 1031 | QTPNSKLVNPKSEDAAGAEGETPHMLRPHVFMPEVTPDMDYLPVFNQGIINPMLS | 1090 |
| DB | 1081 | QTPNSKLVNPKSEDAAGAEGETPHMLRPHVFMPEVTPDMDYLPVFNQGIINPMLS | 1140 |

RESULT 3

| | | | |
|--------|---|------|---------|
| Q6ZM25 | PRELIMINARY; | PRT; | 933 AA. |
| AC | Q6ZM25; | | |
| DT | 05-JUL-2004 (TReMBLrel. 27, Created) | | |
| DT | 05-JUL-2004 (TReMBLrel. 27, Last sequence update) | | |
| DT | 05-JUL-2004 (TReMBLrel. 27, Last annotation update) | | |
| DE | Hypothetical protein FLJ16561. | | |
| OS | Homo sapiens (Human) | | |
| OC | Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; | | |
| OC | Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo. | | |
| OC | NCBI_TaxID=9606; | | |
| RN | [1] | | |
| RP | SEQUENCE FROM N.A. | | |
| RC | TISSUE=Synovial membrane tissue; | | |
| RA | Ota T., Nakagawa S., Senoh A., Mizuguchi H., Inagaki H., Sugiyama T., | | |
| RA | Irie R., Otaki T., Sato H., Wakamatsu A., Ishii S., Yamamoto J., | | |
| RA | Isono Y., Kawai-Hio Y., Saito K., Nishikawa T., Kimura K., | | |
| RA | Yamaehita H., Matsuo K., Nakamura Y., Sekine M., Kikuchi H., Kanda K., | | |
| RA | Wagatsuma M., Murakawa K., Kaneshori K., Takahashi-Fujii A., Oshima A., | | |
| RA | Sugiyama A., Kawakami B., Suzuki Y., Sugano S., Nagahara K., | | |
| RA | Masuko Y., Nagai K., Isogai T., | | |
| RL | Submitted (MAR-2004) to the EMBL/GenBank/DBJ databases. | | |
| DR | ENBL: AK131434; BAD18580.1; - | | |
| DR | InterPro: IPR000585; Hemopexin. | | |
| DR | InterPro: IPR001212; Somatomedin B. | | |
| DR | Pfam: PF00045; Hemopexin; 2. | | |

| | | | |
|--|---|--|-----|
| DR | Pfam: PF01033; Somatomedin B; 1. | | |
| DR | PRINTS; P000022; SOMATOMEDINB. | | |
| DR | SMART; SM00120; HX; 2. | | |
| DR | SMART; SM00201; SO; 1. | | |
| DR | PROSITE; PS00024; HEMOPEXIN; UNKNOWN_1. | | |
| DR | PROSITE; PS00524; SOMATOMEDIN B; 1. | | |
| SQ | SEQUENCE 933 AA; 102512 MW; B883773C5BC21A23 CRC64; | | |
| Query Match 53.7%; Score 3128.9; DB 2; Length 933; | | | |
| Best Local Similarity 54.2%; Pred. No. 1.1e-53; | | | |
| Matches 618; Conservative 0; Mismatches 1; Indels 521; Gaps 3; | | | |
| QY | 1 | MAWKTLPIYLLLLSVFVIQVSSQDLSSCAGCGSGYSDATCNDYNCQHMECCPDF | 60 |
| DB | 1 | MAWKTLPIYLLLLSVFVIQVSSQ | 25 |
| QY | 61 | KRVCTAELSCGKCFESFERGRCDDCAQCKYKDCPDYESFCA | 105 |
| DB | 26 | -----ELSCGKCFESFERGRCDDCAQCKYKDCPDYESFCAEVHNTSPSSKKAP | 79 |
| QY | 106 | -----BEHSVSENQESSSSSSSSSSSSSTIW | 130 |
| DB | 80 | PPSGASQTIKSTTKRSPKPNKKTKKIVIESEBITBEHSVSENQESSSSSSSSSSSTIR | 139 |
| QY | 131 | KIKSKKSAANRELOKLVKDNKNKRTKKKPTKPPVVDVDEAGSLDNGDFKVTTPDTST | 190 |
| DB | 140 | KIKSKKSAANRELOKLVKDNKNKRTKKKPTKPPVVDVDEAGSLDNGDFKVTTPDTST | 199 |
| QY | 191 | TOHKNVSTSPKIIITAKPINRPSLPNSDTSKETSITVKNKETTVETKETTITNKQISTDG | 250 |
| DB | 200 | TOHKNVSTSPKIIITAKPINRPSLPNSDTSKETSITVKNKETTVETKETTITNKQISTDG | 259 |
| QY | 251 | KEKTTSAKETQSIEKTSKADLAPTSKVLAKPTKAEITTKGPAITTPKEPTTPPKBPAS | 310 |
| DB | 260 | KEKTTSAKETQSIEKTSKADLAPTSKVLAKPTKAEITTKGPAITTPKEPTTPPKBPAS | 319 |
| QY | 311 | TTPEKPTPTTIKSAPTTPKBPATTTKSAPTTPKBPATTTKBPATTTKBPATTTTKEP | 370 |
| DB | 320 | TTPEKPTPTTIKSAPTTPKBPATTTKSAPTTPKBPATTTKBPATTTKBPATTTKEP | 370 |
| QY | 371 | APTTTKSAPTTPKBPATTTKBPATTTKBPATTTKBPATTTKBPATTTKBPATTTKBP | 430 |
| DB | 371 | ----- | 370 |
| QY | 431 | EPAPAPKAPKAPATTTKEBPATTPKBPATTTKBPSPITTKBPATTTKSAPTITKEBPAPT | 490 |
| DB | 371 | ----- | 370 |
| QY | 491 | TTKSAPTTPKBPSPITTKBPATTPKBPATTPKBPATTPKBPATTTKBPATTTTCKP | 550 |
| DB | 371 | ----- | 370 |
| QY | 551 | APTAPKEBPATTPKBPATTPKBPATTPKBPATTPKBPATTPKBPATTPKBPATTTT | 610 |
| DB | 371 | ----- | 370 |
| QY | 611 | PEEPAPTPKAAAPNTKBPATTPKBPATTPKBPATTPKBPATTPKBPATTTTKEP | 670 |
| DB | 371 | ----- | 370 |
| QY | 671 | APTTPKAPKAPKAPATTTKEBPATTPKBPATTPKBPATTPKBPATTPKBPATTTT | 730 |
| DB | 371 | ----- | 370 |
| QY | 731 | TAPTITKEBPATTPKBPATTPKBPATTPKBPATTPKBPATTPKBPATTPKBPATTPK | 790 |
| DB | 371 | ----- | 370 |
| QY | 791 | KPAPTTPETPTTSEVSTPTTKEPTTIHKSPDESTPELSAEPPTKALENSPKBPVPT | 850 |
| DB | 371 | ----- | 429 |
| QY | 851 | TKTPAATKPEMTTAKDKTTERDLRTPTTAAAPKMTKETATTTKEKTESKITATTQV | 910 |

CC bronchus, cervix and gall bladder.
 CC -!- PTM: All cysteine residues are involved in intrachain or
 CC interchain disulfide bonds (By similarity).
 CC -!- POLYMORPHISM: The number of repeats is highly polymorphic and
 CC varies among different alleles.
 CC -!- SIMILARITY: Contains 1 C-terminal cysteine knot-like (CTCK) domain.
 CC -!- SIMILARITY: Contains 1 TIL (Trypsin inhibitory-like) domain.
 CC -!- SIMILARITY: Contains 2 WFPC domains.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See [http://www.ebi-sib.ch/](http://www.ebi-sib.ch/announce/)
 CC or send an email to license@sib-sib.ch).
 CC -----
 CC EMBL; L21998; AAB95295.1; -
 CC EMBL; M74027; AAB59875.1; -
 CC EMBL; M94131; AAB59163.1; -
 CC EMBL; M94132; AAB59164.1; -
 CC PIR; A43963; A43932. -
 CC Genew; HGNC:7512; MUC2.
 CC MIM; 158370; -
 CC InterPro; IPR002919; Cys-rich_TIL.
 CC InterPro; IPR006208; Cys_knot.
 CC InterPro; IPR006207; Cys_knot_C.
 CC InterPro; IPR006209; EGF-like.
 CC InterPro; IPR001007; WFC C.
 CC InterPro; IPR001846; WFC D.
 CC Pfam; PF00007; Cys_knot; 1.
 CC Pfam; PF01826; TIL; 1.
 CC Pfam; PF00093; VWC; 1.
 CC Pfam; PF00094; VWD; 4.
 CC SMART; SM00041; CT; 1.
 CC SMART; SM00214; VWC; 2.
 CC SMART; SM00216; VWD; 4.
 CC PROSITE; PS01185; CTCK_1; 1.
 CC PROSITE; PS01225; EGF_1; 1.
 CC PROSITE; PS00222; EGF_1; UNKNOWN_1.
 CC PROSITE; PS01208; WFC_1; 2.
 CC PROSITE; PS0184; WFC_2; 2.
 CC Glycoprotein; Repeat; Signal.
 KW SIGNAL 1 20
 FT CHAIN 21 5179
 FT DOMAIN 1401 1747
 FT REPEAT 1401 1416
 FT REPEAT 1417 1432
 FT REPEAT 1433 1448
 FT REPEAT 1449 1464
 FT REPEAT 1465 1471
 FT REPEAT 1472 1478
 FT REPEAT 1479 1494
 FT REPEAT 1495 1517
 FT REPEAT 1518 1533
 FT REPEAT 1534 1556
 FT REPEAT 1557 1572
 FT REPEAT 1573 1596
 FT REPEAT 1597 1612
 FT REPEAT 1613 1635
 FT REPEAT 1636 1651
 FT REPEAT 1652 1675
 FT REPEAT 1676 1683
 FT REPEAT 1684 1699
 FT REPEAT 1700 1715
 FT REPEAT 1716 1731
 FT REPEAT 1732 1747
 FT DOMAIN 4815 4886
 FT DOMAIN 4924 4991
 FT DOMAIN 5075 5160
 FT DISULFID 5075 5122
 FT DISULFID 5089 5136
 FT DISULFID 5098 5152

FT DISULFID 5102 5154 By similarity.
 FT DISULFID ? 5159 By similarity.
 FT CARBOHYD 163 163 N-linked (GlcNAc. . .) (Potential).
 FT CARBOHYD 423 423 N-linked (GlcNAc. . .) (Potential).
 FT CARBOHYD 670 670 N-linked (GlcNAc. . .) (Potential).
 FT CARBOHYD 770 770 N-linked (GlcNAc. . .) (Potential).
 FT CARBOHYD 894 894 N-linked (GlcNAc. . .) (Potential).
 FT CARBOHYD 1139 1139 N-linked (GlcNAc. . .) (Potential).
 FT CARBOHYD 1154 1154 N-linked (GlcNAc. . .) (Potential).
 FT CARBOHYD 1215 1215 N-linked (GlcNAc. . .) (Potential).
 FT CARBOHYD 1230 1230 N-linked (GlcNAc. . .) (Potential).
 FT CARBOHYD 1245 1245 N-linked (GlcNAc. . .) (Potential).
 FT CARBOHYD 1787 1787 N-linked (GlcNAc. . .) (Potential).
 FT CARBOHYD 1820 1820 N-linked (GlcNAc. . .) (Potential).
 FT CARBOHYD 4339 4339 N-linked (GlcNAc. . .) (Potential).
 FT CARBOHYD 4351 4351 N-linked (GlcNAc. . .) (Potential).
 FT CARBOHYD 4362 4362 N-linked (GlcNAc. . .) (Potential).
 FT CARBOHYD 4373 4373 N-linked (GlcNAc. . .) (Potential).
 FT CARBOHYD 4422 4422 N-linked (GlcNAc. . .) (Potential).
 FT CARBOHYD 4438 4438 N-linked (GlcNAc. . .) (Potential).
 FT CARBOHYD 4502 4502 N-linked (GlcNAc. . .) (Potential).
 FT CARBOHYD 4616 4616 N-linked (GlcNAc. . .) (Potential).
 FT CARBOHYD 4627 4627 N-linked (GlcNAc. . .) (Potential).
 FT CARBOHYD 4752 4752 N-linked (GlcNAc. . .) (Potential).
 FT CARBOHYD 4787 4787 N-linked (GlcNAc. . .) (Potential).
 FT CARBOHYD 4881 4881 N-linked (GlcNAc. . .) (Potential).
 FT CARBOHYD 4888 4888 N-linked (GlcNAc. . .) (Potential).
 FT CARBOHYD 4955 4955 N-linked (GlcNAc. . .) (Potential).
 FT CARBOHYD 4970 4970 N-linked (GlcNAc. . .) (Potential).
 FT CARBOHYD 5019 5019 N-linked (GlcNAc. . .) (Potential).
 FT CARBOHYD 5038 5038 N-linked (GlcNAc. . .) (Potential).
 FT CARBOHYD 5069 5069 N-linked (GlcNAc. . .) (Potential).
 FT CONFLICT 1351 1351 H -> L (in Ref. 3).
 FT CONFLICT 1412 1412 T -> S (in Ref. 3).
 FT CONFLICT 1449 1449 L -> P (in Ref. 3).
 FT CONFLICT 1504 1504 M -> P (in Ref. 3).
 FT CONFLICT 4192 4192 G -> S (in Ref. 2).
 SQ SEQUENCE 5179 AA; 540295 MW; 85CD7571FB9A5663 CRC64;
 Query Match 23.9%; Score 1394.5; DB 1; Length 5179;
 Best Local Similarity 20.2%; Pred. No. 6, 4e-19;
 Matches 449; Conservative 88; Mismatches 95; Indels 1195; Gaps 74;
 QY 6 LPIYLLLSVFVIQVSSQLSSCAGRCG-----EG-----YSR 40
 DB 502 LQVQLAPVQLFVTLDAQSQ--GQVQLGCGNFGLEGDDFKTAGSLVATGAGFANTWKA 559
 QY 41 DATCN-----CDY--- 48
 DB 560 QSTCHDKLWLDPPCLNTESSANYAHWCSSLLKKTETPFGRCHSAVDPAEYKRYKCYDTC 619
 QY 49 NCQHYMEC-----CPDFKRVCTA-----ELSK 71
 DB 620 NCQNNEDCLCAALSSVYARACTAKGVLMGWRHVCKNDVGCPCNSQVFLYLLTTQQTCTCR 679
 QY 72 -----GRCFESF-----BGR-----BDC-----DAQCKYDK 95
 DB 680 SLSEADSHCLGFPAPVDGCGCPDHTFLDEKGRVPLAKGCVHRGLYLEAGVWVQRER 739
 QY 96 C----- 96
 DB 740 CVCRDGRLHCRQIRLIGQSTAPKIHMDCSNLATATSKPRALSQTLAAGYHTECVSG 799
 QY 97 --CPD----- 99
 DB 800 CVCDFGLMDDRGGCVVEKEPCVHNDLYSSGAKIKVDCTCTCKGRWVQTQVCHGT 859
 QY 100 -----YESFCAEHS-----VSENQSSSSSSSSSS 126
 DB 860 CSYIGSGHYITFDGKYDFDGHCSYVAVQDYCGQSSSLGSPSIITENVPCGTTGTCCKA 919
 QY 127 STTWIKSSKNSAANRELQKKLVKDNKK----- 155

Db 920 IKIFMGRT-----ELKLEDKHRVVIQRDEGHVAVYTRVGGYLVESSTGII 967
QY 156 ---NRT---KKKPTPKPPVDEAGSL-----DNGDFKV-----183
Db 968 VIWDEKTTTFFIKLAPSYKGT-----CGLCGNFDRSNNDFTRDHMVVSSSELDGNSWK 1022
QY 184 ---TTPDSTTQ-----HNKYSTSP-----200
Db 1023 EAPTCDVSTNPECSLNPHRHSWAQKQCSILKSSVFSICHSKDPKPYEACVHDSCSC 1082
QY 201 ---TAKPIN-----KI 202
Db 1083 DTGGDCEFCFSAVSAQCTKREGACVFWRTPOLCPIFCDYNNPPECEMHWPEPCGRNSF 1142
QY 203 ---TAKPIN-----PRPSL 214
Db 1143 ETCRTINGIHSNISVYLEGCVPRCPKDRPIVEEDLKKCVADKCGCVVEDTHYPGASV 1202
QY 215 P-----PNSDTSKE-----TSLTVNK 230
Db 1203 PTEETCKSCVCTNSQWCRPEEGKILNOTDGFACYWEICGPNGTVEKHFNICAITRP 1262
QY 231 ET-----TVETKETITNKQTSIDGKEKTTSAK-----258
Db 1263 SLTTFITITLPTTPTSTFTTTTTTPTSTSVLSTPKLCLLMSDWINEDHPSSGSDGD 1322
QY 259 -----ETOSIEKTSKD-----270
Db 1323 REFFDGVCGAPEDIECRSVKDHLSLEHQGQKQVQDVSVGFICKNEQFNGPGLCYDY 1382
QY 271 -----LAPTSKVLAKPTKPAETTTKGPALITPKBPTTTPKPSASTTKBPT 317
Db 1383 KIRVNCWPMDCIITPSPPTTTPSPPTTTPPTTTPSPPTTTPPTTTPSPPTTTPSPPI 1442
QY 318 PTTIKSAPTKE-----PAPTTSKAPTTPKEAPTTPKEAPTTPKEAPTTPKE 369
Db 1443 TTTTTLPTTTPSPPTTTPSPPTTTPSPPTTTPSPPTTTPSPPTTTPPTTTPSPPT 1501
QY 370 P-----APTTSKAP-----TTPKEPA-----PTTP 390
Db 1502 PWTMTPTTTPASTTTLPTTTPSPPTTTPPTTTPPTTTPSPPTTTPPTTTPPTTTPPT 1561
QY 391 KKEAPTTPKEPA-TTPKEPTTTPKEAPTTPKEAPTTPKEAPTTPKEAPTTPKEAPTTPKEPA 449
Db 1562 PSPPPTTTPPTTTPSPPTTTPSPPTTTPPTTTPSPPTTTPPTTTPPTTTPPTTTPSP 1621
QY 450 PTTTKEP-----APTTPKPSPTTPKEAPTTPKSAPT-----TTPKEAPTTPK 493
Db 1622 TTTPTTPTTTPPTTTPSPPTTTPPTTTPPTTTPPTTTPPTTTPPTTTPPTTTPPT 1681
QY 494 SAP-----TTPKPSPTTTPKEAPTTPKEAPTTPKKEAPTTPKEAPTTPKEAPTTPKE-- 542
Db 1682 SSPITTTSPPTTTPPTTTPSPPTTTPPTTTPPTTTPPTTTPPTTTPPTTTPPTTTP 1741
QY 543 -----APTTPKPAAPTKEAPTTP-----KETAPTTPKLT-----576
Db 1742 TMTTLPPTTSPPTTTPPTTTPPTTTPPTTTPPTTTPPTTTPPTTTPPTTTPPTTTP 1799
QY 577 -----576
Db 1800 HKPGGDTLIGDVGFGWAANISCRATWVDPVIGOLGTVVCDVSUGLICKNEDKXPGG 1859
QY 577 -----TTPKEAPTTPKEAPTTPKEAPTTPKEAPTTPKEAPTTPKEAPTTPKEAPTTP 608
Db 1860 VIPMAFCLNLYEINVOCCBCVQPTTMTTNTTENPTTPTTPTTPTTPTTPTTPTTPT 1919
QY 609 -----TTPKEAPTTP-----TPKAAAPTTP-----KE 629
Db 1920 TTTPTTPTTPTTPTTPTTPTTPTTPTTPTTPTTPTTPTTPTTPTTPTTPTTPTTPTTPT 1979
QY 630 PAPTTPKEP-----APTTPKEAPTTPKEAPTTPKEAPTTPKEAPTTPKEAPTTPKEAPT 683
Db 1980 PTPPTGQTPTTPTTPTTPTTPTTPTTPTTPTTPTTPTTPTTPTTPTTPTTPTTPTTPT 2039

RESULT 7
Q9VR49
ID Q9VR49 PRELIMINARY; PRT; 1225 AA.
AC Q9VR49;
DT 01-MAY-2000 (TREMBlrel. 13, Created)
DT 01-JUN-2003 (TREMBlrel. 24, Last sequence update)
DT 01-MAR-2004 (TREMBlrel. 26, Last annotation update)
DE CG3047-PA.
GN Name=Sgs1; ORFName=CG3047;
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_taxid=7227;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Celisner S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanosides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA George R.A., Lewis S.E., Richards J., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Vandeell M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.H., Blazek J.R., Champe M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Gabor G.L.,

684 APTTKEP-----TSTSDKPAPT-----TPKG 706
Db 2040 TTTTPT 2099
QY 707 T-----APTTPKEAP-----TTPKEAPT-----726
Db 2100 TQPT 2159
QY 727 -----TPKGT-APT-----LKEPAPTTPKPAKELAPTTPKGP-----760
Db 2160 VTPPT 2219
QY 761 ---TSTSDKPA-----TTPKETAPTTPKEAP-----TTPKKEAP 794
Db 2220 TPT 2279
QY 795 T-----TPETPP-----PTTSEVS-----TPTTKEPTTTHKS 822
Db 2280 TPTGQTPT 2336
QY 823 PDESTPELSAETPKALENSKEP-----GVPT-TKTPAAKPEMTTAKOKTTE 871
Db 2337 PTTTPT 2396
QY 872 RDLR---TTP-ETTTAAPKMT---KETATTETKTESKI-----TATTQVST 913
Db 2397 TGTQPT 2456
QY 914 T-----TQDTPPKITT-----LKTTLAPKVVTT-KKT 941
Db 2457 TTVPT 2516
QY 942 ITTTEIMNKPEETAKKDRATNSKATTP-----KPKETKAP-----978
Db 2517 PTTPT 2576
QY 979 -KKEPTSKPKMTMPVRKPKXTTTPKMTSTMPKAPTSRIABAMLQTTT-----1027
Db 2577 TPTPTGQTPT 2628
QY 1028 ---RPNQTPNSKLVENPKSEDAG-----GAEGTPEMLRAPHV 1063
Db 2629 TQPT 2688
QY 1064 FMEVTP 1070
Db 2689 VTPPTPT 2695

RA Abril J.F., Agbayani A., An H.J., Andrews-Pfannkuch C., Baldwin D.,
RA Balow R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
RA Burtis K.C., Busan D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Douc L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferriera S., Fleischmann W.,
RA Foster C., Gabriellian A.E., Garg N.S., Gelbart W.M., Glaser K.,
RA Glöckel A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.H., Ibegwam C.,
RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Maitel B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merklov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D., Scheeler F., Shen H.,
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E., Wang X.,
RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.Y., Wasserman D.A., Weinstein G.M., Weissbach J.,
RA Williams S.M., Woodgett, Worley K.C., Wu D., Yang S., Yao Q.A., Ye J.,
RA Yeh R.F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C., Smith H.O.,
RT "The genome sequence of *Drosophila melanogaster*."
RL Science 287:2185-2195(2000).
RN [2]
RX SEQUENCE FROM N.A.
RX MEDLINE=22426065; PubMed=12537568;
RA Celniker S.E., Wheeler D.A., Krommiller B., Carlson J.W., Halpern A.,
RA Patel S., Adams M., Champe M., Dugan S.P., Frise E., Hodgson A.,
RA George R.A., Hoskins R.A., Lavery T., Muzny D.M., Nelson C.R.,
RA Pacleb J.M., Park S., Pfeiffer B.D., Richards S., Sodergren E.J.,
RA Svirskas R., Taber P.E., Wan K., Stapleton M., Sutton G.G., Venter C.,
RA Weinstein G., Scherer S.E., Myers E.W., Gibbs R.A., Rubin G.M.,
RT "Finishing a whole-genome shotgun: release 3 of the *Drosophila*
RT melanogaster euchromatic genome sequence."
RT Genome Biol. 3:RESEARCH0079-RESEARCH0079(2002).
RN [3]
RX SEQUENCE FROM N.A.
RX MEDLINE=22426070; PubMed=12537573;
RA Kaminker J.S., Bergman C.M., Krommiller B., Carlson J., Svirskas R.,
RA Patel S., Frise E., Wheeler D.A., Lewis S.E., Rubin G.M.,
RA Ashburner M., Celniker S.E.,
RT "The transposable elements of the *Drosophila melanogaster* euchromatin:
RT a genomics perspective."
RL Genome Biol. 3:RESEARCH0084-RESEARCH0084(2002).
RN [4]
RX SEQUENCE FROM N.A.
RX MEDLINE=22426069; PubMed=12537572;
RA Mira S., Crosby M.A., Mungall C.J., Matthews B.B., Campbell K.S.,
RA Hradecky P., Huang Y., Kaminker J.S., Millburn G.H., Prochuk S.E.,
RA Smith C.D., Tupy J.L., Whitfield E.J., Bayraktaroglu L., Berman B.P.,
RA Bettencourt B.R., Celniker S.E., de Grey A.D., Drysdale R.A.,
RA Harris N.L., Richter J., Russo S., Schroeder A.J., Shu S.Q.,
RA Stapleton M., Yamada C., Ashburner M., Gelbart W.M., Rubin G.M.,
RA Lewis S.E.,
RT "Annotation of the *Drosophila melanogaster* euchromatic genome: a
RT systematic review."
RL Genome Biol. 3:RESEARCH0083-RESEARCH0083(2002).
RN [5]
RX SEQUENCE FROM N.A.
RG FLYBASE;
RN Submitted (SEP-2002) to the EMBL/GenBank/DBJ databases.
RN [6]
RX SEQUENCE FROM N.A.
RN FLYBASE;

RL Submitted (MAR-2004) to the EMBL/GenBank/DBJ databases.
DR EMBL: AF003575; AAF50957.3; --
DR FLYBASE: FBgn003372; Sgsl.
SQ SEQUENCE 1225 AA; 127735 MW; A99AF9D4404C79F4 CRC64;
Query Match 22.7%; Score 1321.9; DB 2; Length 1225;
Best Local Similarity 26.8%; Pred. No. 2.7e-18;
Matches 358; Conservative 118; Mismatches 419; Indels 441; Gaps 51;
QY 10 LLLLLLVFIQV-----SSDLSACGCGEGYSDATCNDYNOCHMECCPDF 60
DB 6 IFLTIVLIIQVKNVANYVDWDSMD-----GFSEIIPCGGDIY-----PD- 50
QY 61 KRVCYAEISCKGCFESFERGECDCDAQCKYKCCPDYEFCAEHSVSENOSSSS 120
DB 51 -----PVQPCDTD----- 58
QY 121 SSSSSSTIWKISKNSNAARELOKLLK-----VKDNKKNTKKK-TPK 165
DB 59 ---SNPTTKPKQTKRPKSTRRTTKRPRKTKTKTKTKTKTKTKTKTKTKTK 115
QY 166 PPVDEAGSGLDNGDFKVTTPDT-----STQHNVKVS-----TSPKITTAKP 207
DB 116 TP-----DTDSPIITGAECTCSDRITTSSTDTSDTDTTNTDWTPLCTDTP 166
QY 208 INRPSPNPNSDTSKETSIVNKEITVE-----TKETTTNKOTSDGKKT--T 255
DB 167 CT-----CSSESSTAIPSPCIDTSTVIPSCTQESTTPTTCSTGGTQTTPT 216
QY 256 SAKETQSIEKTSKDLAPTSKVLAKTPKAET--TKGPALTTPKEPTTPKEPASTTP 313
DB 217 CQQT-----TTPRSTTTTSRPTTTPRSTTTTTSRPTTTP 256
QY 314 KEPTPTTIKSAPTPKEPAPTTTKSAP--TPKEPAPTTKEPAPTTKEPAPTTKEPA 371
DB 257 RSTTTTTRPTTTPRPTTTCAPTITTPRSTTTTTSRPTTTPRSTTTTSCSPT 316
QY 372 PTTTKSAPTT-PKEPAPTT-----KKPAPTTTPE-----PAPTTTPE--- 408
DB 317 RTPRSTTTTTSRPTTTPRPTTTPRPTTTPRPTTTPRPTTTPRPTTTPRPTTTP 376
QY 409 -----PTPTTPEPAPT-TKEPAPTTPEPAPTAPKAPPTTPE-----PAPTTPK 454
DB 377 TCTSGPTTTTTPRSTTTTSGPTTTPRSTTTTSGPTTTPRSTTTTSGPTTTPR 436
QY 455 EPAPTTKEPSTTPKEPAPTTTKSAPTTKEPAPTTKSAP--TTPKEPSTTTKEAP 512
DB 437 STTTTSGPTTTPRSTTTTSGPTTTPRSTTTTSGPTTTPRSTTTTSGPTTTPR 496
QY 513 TTPKE-----PAPTTTPE-----PAPTTTPE-----PAPTTTPEPTTK 548
DB 497 TTPRSTTTTSGPTTTPRSTTTTSGPTTTPRSTTTTSGPTTTPRSTTTTSGPT 556
QY 549 KPAPTPKE-----PAPTTPEAPTTPKLTPEKLTPEKLTPEKLTPEKLTPEELAP 600
DB 557 GPTTTTTPRSTTTTSGPTTTPRSTTTTSGPTTTPRSTTTTSGPTTTPRSTTT 616
QY 601 TPEEPTTTPPE-----PAPTTKAAAPTKEPAPTTTPE-----PAPTTPK 644
DB 617 TSTSGPTTTPRSTTTTSGPTTTPRSTTTTSGPTTTPRSTTTTSGPTTTPRST 676
QY 645 E-----PAPTTTPEAP-----TTPKGTAPTLKEPAPTTTPE----- 677
DB 677 STTTTSGPTTTPRSTTTTSGPTTTPRSTTTTSGPTTTPRSTTTTSGPTTTPR 736
QY 678 PAPKELAPTTTKEPTSTT-----SDKPAPTTPKGTAPTTKEPAPTTTPE----- 722
DB 737 TTPRSTTTTSGPTTTPRSTTTTSGPTTTPRSTTTTSGPTTTPRSTTTTSGPT 796
QY 723 -PAPTTTKEPAPTTTKEPAPTTTPE-----PAPKELAPTTTKEPAPTTT 764
DB 797 GPTTTTTPRSTTTTSGPTTTPRSTTTTSGPTTTPRSTTTTSGPTTTPRSTTT 856

```

QY 765 --SDKPAPTPKETAPTTPKEAPTTPKX-----PAPTTPET-----PPPTTSE 806
Db 857 TGTSCPTTTTPRSTTTCTGTTTTPRSTTTCTGTTTTPRSTTTCTGTTTTPRSTTTCTGTTTTPR 916
QY 807 VSTPTTKETPTTHKSPDESTELSAETPKALENSPKGVPVTKTTPAAKPEMTTAK 866
Db 917 STTTTCTSGPTT--TTPRSTTKTSCAPT-----TTPRSTTTTTSRPTTTTPRSTTTT- 969
QY 867 DKTTTDLRLTTPETTTAAKPMKTKETATTEKTESKITATTQVTTSTTTQDTPPKITL 926
Db 970 --TTSRPTTTTPRSTT--TPTSRPTTTTPRSTT-----TTSRPTTTTPRSTT----- 1015
QY 927 KTTTLAPKVTTHKTLITTEIMNKBEETAKPKDRATNSKATPKPKQKTKAPK----- 979
Db 1016 KSTCAPTTTTPRSTTTT-----TSRPTTTTTPRSTTTTTPRSTTTTTPRSTTTTPTCT 1067
QY 980 -KPTSTKPKTMRVRKPKTTPRKMVT-----STMPELNPTSRIAEAMLQTTTR----- 1028
Db 1068 SRPTTTTPRSTTTTTSRPTTTTPRSTTTPCPTTTPSASPTR-----TTPTRPCPCH 1120
QY 1029 -----PNQT-----PNSKLVNPKSEDAGGAGETPHMLLRPHVFM 1065
Db 1121 POPPYQIPWWSWYENPTYPNPWPQPN-----PNSKLVNPKSEDAGGAGETPHMLLRPHVFM 1149
QY 1066 PEVTPDMDYLPRVNO 1081
Db 1150 -PVFPQWPCMPYPNQ 1164

RESULT 8
ID Q7KTF6 PRELIMINARY; PRT; 1761 AA.
AC Q7KTF6
DT 05-JUL-2004 (TrEMBLrel. 27, Created)
DT 05-JUL-2004 (TrEMBLrel. 27, Last sequence update)
DE CG33300-PA (Fragment).
DE CG33300-PA (Fragment).
GN ORFNames=CG33300;
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxId=7227;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=20196006; PubMed=10731132;
RA Adams W.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galie R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Vandeil M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.H., Blasej R.G., Champe M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Gabor G.L.,
RA Abril J.F., Agbayani A., An H.J., Andrews-Pfannkoch C., Baldwin D.,
RA Ballwe R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brotter P.,
RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablo B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
RA Glöck A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.H., Ibegwam C.,
RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai X.,
RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nussekern D.R., Paclele J.M.,
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,

```

```

RA Reinert K., Remington K., Saunders R.D., Scheeler F., Shen H.,
RA Shue B.C., Siden-Klamos I., Simpson M., Skupski M.P., Smith T.,
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.Y., Wasserman D.A., Weinstock G.M., Weissenbach J.,
RA Williams S.M., Woodage, Worley K.C., Wu D., Yang S., Yao Q.A., Ye J.,
RA Yeh R.F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RA "The genome sequence of Drosophila melanogaster.";
RL Science 287:2185-2195(2000).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=22426065; PubMed=12537568;
RA Celniker S.E., Wheeler D.A., Kronmiller B., Carlson J.W., Halpern A.,
RA Patel S., Adams M., Champe M., Dugan S.P., Frise E., Hodgson A.,
RA George R.A., Hoskins R.A., Laverty T., Muzny D.M., Nelson C.R.,
RA Paclele J.M., Park S., Pfeiffer B.D., Richards S., Sodergren E.J.,
RA Svirskas R., Tabor P.E., Wan K., Stapleton M., Sutton G.G., Venter C.,
RA Weinstock G., Scherer S.E., Myers E.W., Gibbs R.A., Rubin G.M.;
RA "Finishing a whole-genome shotgun: release 3 of the Drosophila
RA melanogaster euchromatic genome sequence.";
RL Genome Biol. 3:RESEARCH0079-RESEARCH0079(2002).
RN [3]
RP SEQUENCE FROM N.A.
RX MEDLINE=22426070; PubMed=12537573;
RA Kaminker J.S., Bergman C.M., Kronmiller B., Carlson J., Svirskas R.,
RA Patel S., Frise E., Wheeler D.A., Lewis S.E., Rubin G.M.,
RA Ashburner M., Celniker S.E.;
RA "The transposable elements of the Drosophila melanogaster euchromatin:
RA a genomics perspective.";
RL Genome Biol. 3:RESEARCH0084-RESEARCH0084(2002).
RN [4]
RP SEQUENCE FROM N.A.
RX MEDLINE=22426069; PubMed=12537572;
RA Misra S., Crosby M.A., Murgall C.J., Matthews B.B., Campbell K.S.,
RA Hradecky P., Huang Y., Kaminker J.S., Millburn G.H., Prochnik S.E.,
RA Smith C.D., Tupy J.L., Whitfield E.J., Bayraktaroglu L., Berman B.P.,
RA Bettencourt B.R., Celniker S.E., de Grey A.D., Drysdale R.A.,
RA Harris N.L., Richter J., Russo S., Schroeder A.J., Shu S.Q.,
RA Stapleton M., Yamada C., Ashburner M., Gelbart W.M., Rubin G.M.,
RA Lewis S.E.;
RA "Annotation of the Drosophila melanogaster euchromatic genome: a
RA systematic review.";
RL Genome Biol. 3:RESEARCH0083-RESEARCH0083(2002).
RN [5]
RP SEQUENCE FROM N.A.
RX FLYBASE;
RL Submitted (SEP-2002) to the EMBL/GenBank/DBJ databases.
RN [6]
RP SEQUENCE FROM N.A.
RX FLYBASE;
RL Submitted (MAR-2004) to the EMBL/GenBank/DBJ databases.
DR EMBL; AE003626; AAS64673.1; -
DR InterPro; IPR011009; Kinase_like.
FT NON TER
SQ SEQUENCE 1761 AA; 197521 MW; 62A5E16E6241B3F4 CRC64;
Query Match 22.6%; Score 1314.1; DB 2; Length 1761;
Best Local Similarity 24.8%; Pred. No. 6e-18;
Matches 450; Conservative 117; Mismatches 376; Indels 869; Gaps 84;
QY 74 CF-----ESFPERGECDCDCAQCKKYDKC-----CP-----DYESFCAE 106
Db 5 CFIWLALYHSANAALGSEIKFGDESTTESAEKYDYCWINPFLPGCFPGD----- 57
QY 107 EHSVSENBQSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSS 163
Db 58 -----GNNSTTKISTLATTKSTVTTSEETTLTKTKTKTKTKTKTKTKTKTKTKTKTKTK 105
QY 164 PKPVPVDEAGSLDNGFKVTTTDTST-----TQHNKV-----STSPKI 202
Db 106 TE-----KITTPKATIKSTATTATATAPKTEQTLTKTKTKTKTKTKTKTKTKTKTKTK 149

```



```

Db 1033 LKPTREPTTKSTTKTTREPTTKVTTERTREPTTSKTTTHETTAEPATKKTTHETPTT 1092
QY 704 PKGTA-----PTTPKEPAPTTPKEPA-----PTTPKGTAPTTLKEPA----- 740
Db 1093 OKSTTLRTEBPTTKSTAKTTRPTTKRETTERTKEPTTKTTHKTTHEEPTTKTT 1152
QY 741 --PTTPK-----KPAPKELAPITTKGPTSTTSDDKAPATPKETAPTTPKEPA----- 785
Db 1153 HEPPTTKSTTLKPTBE-----PTTRK--TSITKTRBPTTKRKTERTTKBPTTKTTHKT 1207
QY 786 -----PTTPK-----KP-----APTTPETPPPTSEVSTPTTKKEP 816
Db 1208 TEEPTTKNTTKTTHETTKSTTLKPTBPTTKRKTSTTKTTRPTTKRVTTERPTREP 1267
QY 817 ---TTIHKSPDE-----STPELSAETPKALNSPKPGVPTTKTAATKPEMTTAK 866
Db 1268 TTRKTTTHKTEBPTTKKTTTKTTHETPTTK--STTLKPTBPTTKSTTKTTRPTTK 1326
QY 867 DKTERDLR-----PTTPK-----TPETTTAAP----- 885
Db 1327 RVTERTRPTTKRKTTHKTEBPTTKKTTTKTTHETTKKSTTLKPTBPTTKTST 1386
QY 886 -KMKETAT--TTEKTESKITATTQVSTTTQDTTPPKITLKT-----TTLA 932
Db 1387 TKTRPTTKRVTTERTRPTTKT--THKTTBPTTKKTTTKTTHETTKKSTTLK 1444
QY 933 P-----KVTTK-----KITTHIMNKP-----BETAKPKDRATNSKATPK 970
Db 1445 PTEBPTTKSTTKTTRPTTKRVTTERTRPTTKTTHKTEBPTTKTTHETTKTTHET 1504
QY 971 PQ-----KP-----TKAPKKPT----- 982
Db 1505 PTKKSTTLKTEBPTTKSTTKATREPTTKRVTTERTRPTTKTTHKTEBPTTK 1564
QY 983 -----STKKPKTMPVRPKPTTPRKMSTWPELNPTSRABAMLOTT----- 1026
Db 1565 KTTTKKTHETPTTKSTL-----KTEBPTTKSTTKTTRPTTKSTSVKTTADQIT 1620
QY 1027 -----TRPNQTPNSKLVEVNPKESEDAGGAGET-----PHM 1057
Db 1621 KETTAEMSTTNOBTSVETTTNSQNTTSTSTBEOVHHHHHHHHYHKPADJGPSI 1680
QY 1058 LLRPH----- 1069
Db 1681 LPLFDLPPLPLPWPPLPLPEPLPLPLPTALPPLPLPLPLPLPEVNLTAISLPEIS 1740
QY 1070 -PMDYLPRVEN 1080
Db 1741 LPNLPPLPLQLEN 1752

```

RESULT 10
Q7PMD5

ID Q7PMD5 PRELIMINARY; PRO; 3150 AA.
AC Q7PMD5;
DT 01-MAR-2004 (Tremblrel. 26, Created)
DT 01-MAR-2004 (Tremblrel. 26, Last sequence update)
DT 01-MAR-2004 (Tremblrel. 26, Last annotation update)
DE ENSANGP0000004655 (Fragment).
GN Name=ENSANGP00000003651;
OS Anopheles gambiae str. PEST.
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Nematocera; Culicoidea; Anopheles.
OX NCBI_TaxID=180454;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=PEST;
RA Anopheles Genome Sequencing Consortium;
RL Submitted (APR-2003) to the EMBL/GenBank/DBJ databases.
CC -! CAUTION: The sequence shown here is derived from an
CC EMBL/GenBank/DBJ whole genome shotgun (WGS) entry which is
CC preliminary data.

```

DR EMBL; AAAB01008980; EAA13969.2; --
DR GO; GO:0016020; C:membrane; IEA.
DR GO; GO:0004872; F:receptor activity; IEA.
DR GO; GO:0008151; P:cell growth and/or maintenance; IEA.
DR InterPro; IPR006770; OGF:recept.
DR Pfam; PF04680; OGF_r_III; 80.
FT NON_TER 1
FT SEQUENCE 3150 3150
SQ SEQUENCE 3150 AA; 322879 MW; 3C7B3D441C58C839 CRC64;

Query Match 21.3%; Score 1243.6; DB 2; Length 3150;
Best Local Similarity 19.9%; Pred. No. 3e-16;
Matches 402; Conservative 111; Mismatches 397; Indels 1114; Gaps 74;

QY 111 SENQSSSSSSSSSSSSSIWIKSKNSGAANRELOKLVKDNKKKTKKTPKPPVVD 170
Db 870 TESTDTTMSASSTPEPST-----TSGTTR-----TTTPTPTPTDTTMS 907
QY 171 EAGS-----GLDNGDFKVTTPDTSTTQHN----- 194
Db 908 SASSTPEPSTTPGTTTTPRPTSTESTDTTMSASSTPEPSTTPGTTTTPRPTSTESTD 967
QY 195 -----KVSTSPKITTAKPINP-----RPSLP 215
Db 968 TTMSSASSTPEPSTTPGTTTTPRPTSTESTDTTMSASSTPEPSTPKPGTTRTTPRPTT 1027
QY 216 PMSDTSKETSILVNKETTVEKTTTNTKQSTDGKKTSA----- 257
Db 1028 ESTDTTMSASSTESTPTTPGTTTTPRPTSTESTDTTMSASSTPEPSTTPGTTTTPRPT 1086
QY 258 --KETQSSEKTSKADLAP---TSKVLAKPTPKAET-----TTKGPAL 294
Db 1087 RPTSTESTDTTMSASSTPEPSTTPGTTTTPRPTSTESTDTTMSASSTPEPSTTPGTT 1146
QY 295 TTPKSEPTP-----TTPKEPAS----- 310
Db 1147 TTPPTSTESTDTTMSASSTPEPSTTPGTTTTPRPTSTESTDTTMSASSTPEPSTTP 1206
QY 311 ---TTPKEP-----TPTTIKSAPTTP-----KEPA-- 332
Db 1207 GTTTRTTPRPTSTESTDTTMSASSTPEPSTTPGTTTTPRPTSTESTDTTMSASSTPEPST 1266
QY 333 PTTTKSAPTTP-----KEPA 347
Db 1267 PGTTRTTPRPTSTESTDTTMSASSTPEPSTTPGTTTTPRPTSTESTDTTMSASSTPE 1326
QY 348 PTTTKSAPTTPKEAPT-----TTKEPA--PTTTKSAPTTP----- 382
Db 1327 PTTTPTDTRTTPRPTSTESTDTTMSASSTPEPSTTPGTTTTPRPTSTESTDTTMSA 1386
QY 383 KEPA-----TTPKKAPPT-----TTPKEAPTTPKEAPT----- 420
Db 1387 STPEPSTTPGTTTTPRPTSTESTDTTMSASSTPEPSTTPGTTTTPRPTSTESTDTT 1446
QY 421 ---TKEPAP-----TTPKEAPT-----APKKAP-----TTPKEPAP-- 450
Db 1447 MSSASSTPEPSTTPGTTTTPRPTSTESTDTTMSASSTPEPSTTPGTTTTPRPTSTES 1506
QY 451 -----TTPKEAPTTPKPSPTTPKEPA-----PTTT 477
Db 1507 TDTTMSASSTPEPSTTPGTTTTPRPTSTESTDTTMSASSTPEPSTTPGTTTTPRPTST 1565
QY 478 KSAPT-----TKBPA--PTTTKSAPTTP-----KPSPTTTPKEAPTTPKE 517
Db 1566 ESTDTTMSASSTPEPSTTPGTTTTPRPTSTESTDTTMSASSTPEPSTTPGTTTTPR 1625
QY 518 PAP-----TTPKKAP-----TTPKKAP-----TTPK 532
Db 1626 PTTSTESTDTTMSASSTPEPSTTPGTTTTPRPTSTESTDTTMSASSTPEPSTTPGTTT 1685
QY 533 EPAPT-----TTPKEAPTTPKKA-----PTAP-----KEAPT 561
Db 1686 RPTPTDSTMSMSESTPEPSTTPGTTTTPRPTSTESTDTTMSASSTPEPSTTPGTTT 1745

```


RESULT 13
 SLPI_CLOTM STANDARD; PRT; 1664 AA.
 AC Q06852; 1994 (Rel. 29, Created)
 DT 01-JUN-1994 (Rel. 29, Last sequence update)
 DT 05-JUL-2004 (Rel. 44, Last annotation update)
 DE Cell surface glycoprotein 1 precursor (Outer layer protein B) (S-layer
 DE protein 1).
 GN Name=olpB;
 OS Clostridium thermocellum.
 OC Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
 OC Clostridium.
 OX NCBI_TaxID=1515;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=NCIB 10682;
 RX MEDLINE=93209931; PubMed=8456832;
 RA Fujino T., Beguin P., Aubert J.-P.;
 RT "Organization of a Clostridium thermocellum gene cluster encoding the
 RT cellulosomal scaffolding protein CipA and a protein possibly involved
 RT in attachment of the cellulosome to the cell surface.";
 RL J. Bacteriol. 175:1891-1899 (1993).
 CC -! SUBUNIT: Assembled into mono-layered crystalline arrays.
 CC -! SUBCELLULAR LOCATION: Cell wall.
 CC -! SIMILARITY: Contains 4 S-layer homology (SLH) domains.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation.
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See [http://www.isb-sib.ch/](http://www.isb-sib.ch/announce/)
 CC or send an email to license@isb-sib.ch).
 CC
 DR EMBL; X67506; CAA47841.1; -
 DR PIR; T18262; T18262.
 DR InterPro; IPR008965; Cellul_bind.
 DR InterPro; IPR001119; SLH.
 DR Pfam; PF00395; SLH; 3.
 DR PROSITE; PS01072; SLH DOMAIN; 2.
 KW Cell wall; Repeat; S-layer; Signal.
 FT SIGNAL 1 28
 FT CHAIN 29 1664
 FT DOMAIN 36 763
 FT REPEAT 36 191
 FT REPEAT 207 363
 FT REPEAT 409 565
 FT REPEAT 607 763
 FT DOMAIN 771 1377
 FT DOMAIN 1378 1449
 FT DOMAIN 1453 1494
 FT DOMAIN 1495 1565
 FT DOMAIN 1566 1625
 FT DOMAIN 1626 1646
 FT SEQUENCE 1664 AA; 178194 MW, 5F396655BA9FE74B CRC64;
 Query Match 19.3%; Score 1127.2; DB 1; Length 1664;
 Best Local Similarity 21.4%; Pred. No. 2.5e-14;
 Matches 369; Conservative 127; Mismatches 427; Indels 798; Gaps 68;
 QY 4 KTLPIYLLLLSVFVIOQVSSDLSACAGCGEGYSRDATCNCYNCQHYMECCP----- 58
 DB 6 KVLISILLTLL-----IISTSVNKS-----FAEATPSIEMV 37
 QY 59 -----DFKRVCTAELSC-----GRCFESFGRGDC 86
 DB 38 LDKTEVHVGDVITATIKVNNIRKLQYQNLKIPDPVLPVDPVDPATGEEFT----- 87
 QY 87 DAQCKYDKCCP-----DYESFCAEEH-- 108

Db 88 -----DKSMFVNRVLLTNSKYGTPVAGNDIKSGIINFATGYNLNTAYKSSGIDEHTG 140
 QY 109 -----SVSENQESSSSSSSSSSSTTW-----KIKSS 135
 Db 141 IIGBIGFKVLKQNTSIRFDTLSMPGCAISGTSLFDWDAETITGYEVIQPDLLVVAEPL 200
 QY 136 KNGAANRELOK-KLVKVD-----NKN-----RTKKPT----- 163
 Db 201 KDAVALELDKTKVKGDIITAIKIKENKVFAGYQNLKIDFTMLEALETGTGSAIKR 260
 QY 164 -----PKPVDVDEAGSLDN----- 178
 Db 261 TWPVTGTVLQSDNYGKTTAVANDVGAGIINFAEAYSNTKYRETGVAETGIIGKIGPR 320
 QY 179 ---GDFKVTTPDT-----STQHNKV-----STSPK 201
 Db 321 VLKAGSTAIRFEDTTAMPGAIEGTYMFDWYGENIKGYSVQPGIEVAEGEERPEEPT 380
 QY 202 ITTAKTINPRPSLP-----PNS-----DTSK----- 222
 Db 381 VPTETVPDPTVTEBPVPSLPDSVYIMELDKTKVKVGDIITAIKIKENKVFAGYQNL 440
 QY 223 -----ETSLTVNKETTVETKETTNNKOTSTDGKEKTS----- 256
 Db 441 IKYDPTMLEALETGTSAIKATWPTGGTV-----LQSDNYGKTTAVANDVGAGIINFA 495
 QY 257 -----AKETQSIEKTSADLAPTSKVLAKPTKAEETT----- 289
 Db 496 EAVSNLTKYRETVGAETGIGKIGFVLKAGSTAI-----RPEDTAMPGAIEGTYMFD 550
 QY 290 -----KGPAITTP-----KEPTPT-TPKEPASTTPKEPTPT----- 319
 Db 551 WYGENIKGYSVQPGIEVAEGEETPEPTVPTDPTVTEBPVPSLPDSVYIMELD 610
 QY 320 ----- 319
 Db 611 KTKVREGDVIIATIRVNNIKNLAGYQIGIKYDPKVLKLEAFNIETGDPIDEGTWAPVGGTIL 670
 QY 320 -----TIKSAPTTK 329
 Db 671 KNRDYLPTGVAINNVSKGILNFAAYVYVDDYREKGGEDTGIIIGNIFVLKAEDETTIR 730
 QY 330 -----EPATTTKSAPTTKPEPATTKEPATTTP 359
 Db 731 FFELESMPGSDGTWMLDNLNRIISGVYVQAPIKAAAS-----DEPTDTPSDEPTPS 785
 QY 360 KEPAP-----TTTKEPATTTKSAPTTKPEPATTTPKPAITTPKEPATT-----TP 406
 Db 786 DEPTPSDEPTPSDEPTPSDEPTPEEPIPTDTPSDEPTPSDEPTPSDEPTPSDEPTP 845
 QY 407 K-EPTPT-TPKEPATT-TPKEPATTTPKEPATTTPKPA--PTTPKEPATT-TPKEPATT 460
 Db 846 SDEPTPEETPEEPIPTDTPSDEPTPSDEPTPSDEPTPSDEPTPSDEPTPEEPIPTD 905
 QY 461 TKPEPTTPKEPATTTKSAPTTKPEPATTTKSAPTTKPEPSTTTKEPATTTPKEPATT 520
 Db 906 TSDEPTPSDEPTPS-----DEPTPSDEPTPSDEPTPEEPIPTDTPSDEPTPSDEPTP 962
 QY 521 TTPKKA-----PTTPKEPATT-TPKEPATTTKKPAITTPKEPATTTPKE 565
 Db 963 SDEPTPSDEPTPSDEPTPSDEPTPEEPIPTDTPSDEPTPSDEPTPSDEPTPSD--- 1019
 QY 566 TAPTPKLTTPTPKLAITTPKEPATT-----TPELAPTTPEEPTPT-TPKEPATTTPKA 621
 Db 1020 -----EPTPSDEPTPSDEPTPSD-EPTPSDEPTPSDEPTPEEPIPTDTPS 1061
 QY 622 AAPNTKSPAPTTKPEA--PTTPKEPATT-TPKETAPTTKGTAPTTKSPAPTTKPP 678
 Db 1062 DEPTPSDEPTPSDEPTPSDEPTPEEPIPTDTPSDEPTPSDEPTPS--DEP 1119
 QY 679 APKELAPTTKEPT--STTSDKPAITTPKGTAPTTKPEPATTTPKEPATTTPKGTAPTLK 737
 Db 1120 TSPD-EPTPSDEPTPEEPIPTDTPSDEPTPSDEPTPS--DEPTPS--DEPTPSD 1172


```

QY      822 SPDESTPELSAETPKALENSKPEKPGVTTTKTAAKPMVTTAKDKTTERDLRTTPTTT 881
DB      1187 SAATSTTSGSGTTP-----SP-----VPTTSTTSASTTSTTSAPTSTT-----SGPGTT 1232
QY      882 TAAPKMTKATTTTETKTSKTAATTQVTTSTTQ--DTTPKTTTLTKTTLAPKVVTTK 939
DB      1233 PSEVPSTSTSAATSTTSGAPTTRTTSAPTSTSGPGTTPSEVPSTTSTTS-APTST-- 1289
QY      940 KUITTEINNKBEATAKPKDRATNSKATPKQKPKAPKP--TSTKPKMTPVRKPKPT 998
DB      1290 -----TSGPGTTPSEVPSTTSTTSAPTSTT-----TSGPGS 1319
QY      999 TTPPKVKTSTMPELNPTSTSIAPAMLTQTTT 1027
DB      1320 TSPVPTTSTTSA--PTTSTTSASTSTT 1346

RESULT 15
O76894 PRELIMINARY; PRT; 1795 AA.
AC O76894;
DT 01-NOV-1998 (TrEMBLrel. 08, Created)
DT 01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
DT 05-JUL-2004 (TrEMBLrel. 27, Last annotation update)
DE CG14796-PA (EG:5667.1 protein).
GN Name=EG:5667.1; ORFNames=CG14796;
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Ceiniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers J.H., Blazej R.G., Champe M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Gabor G.L.,
RA Abril J.F., Agbayani A., An H.J., Andrews-Pfannkoch C., Baldwin D.,
RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
RA Burtis K.C., Busan D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA de Fabios B., Decher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
RA Foster C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA Hoslin D., Houston K.A., Howland T.J., Wei M.H., Ibegwam C.,
RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacle J.M.,
RA Palazolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D., Scheeler F., Shen H.,
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Svirskaas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.Y., Wasmann D.A., Weinstock G.M., Weissenbach J.,
RA Williams S.M., Woodgett, Worley K.C., Wu D., Yang S., Yao Q.A., Ye J.,
RA Yen R.F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhou S., Zhu S.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT "The genome sequence of Drosophila melanogaster."
RL Science 287:2185-2195(2000).

```

```

RN RP SEQUENCE FROM N.A.
RX MEDLINE=22426065; PubMed=12537568;
RA Ceiniker S.E., Wheeler D.A., Kronmiller B., Carlson J.W., Halpern A.,
RA Patel S., Adams M., Champe M., Dugan S.P., Frise E., Hodgson A.,
RA George R.A., Hoskins R.A., Javerty T., Muzny D.M., Nelson C.R.,
RA Pacle J.M., Park S., Pfeiffer B.D., Richards S., Sodergren E.J.,
RA Svirskaas R., Tabor P.E., Wan K., Stapleton M., Sutton G.G., Venter C.,
RA Weinstock G., Scherer S.E., Myers E.W., Gibbs R.A., Rubin G.M.;
RT "Finishing a whole-genome shotgun: release 3 of the Drosophila
RT melanogaster euchromatic genome sequence."
RL Genome Biol. 3:RESEARCH0079-RESEARCH0079(2002).
RN [3]
RP SEQUENCE FROM N.A.
RX MEDLINE=22426070; PubMed=12537573;
RA Kaminker J.S., Bergman C.M., Kronmiller B., Carlson J., Svirskaas R.,
RA Patel S., Frise E., Wheeler D.A., Lewis S.E., Rubin G.M.,
RA Ashburner M., Ceiniker S.E.;
RT "The transposable elements of the Drosophila melanogaster euchromatin:
RT a genomics perspective."
RL Genome Biol. 3:RESEARCH0084-RESEARCH0084(2002).
RN [4]
RP SEQUENCE FROM N.A.
RX MEDLINE=22426069; PubMed=12537572;
RA Misra S., Crossy M.A., Mungall C.J., Matthews B.B., Campbell K.S.,
RA Hradecky P., Huang Y., Kaminker J.S., Milburn G.H., Prochuk S.E.,
RA Smith C.D., Tupy J.L., Whitfield E.J., Bayraktaroglu L., Berman B.P.,
RA Bettencourt B.R., Ceiniker S.E., de Grey A.D., Drysdale R.A.,
RA Harris N.L., Richter J., Russo S., Schroeder A.J., Shu S.O.,
RA Stapleton M., Yamada C., Ashburner M., Gelbart W.M., Rubin G.M.,
RA Lewis S.E.;
RT "Annotation of the Drosophila melanogaster euchromatic genome: a
RT systematic review."
RL Genome Biol. 3:RESEARCH0083-RESEARCH0083(2002).
RN [5]
RP SEQUENCE FROM N.A.
RX FLYBASE;
RL Submitted (SEP-2002) to the EMBL/GenBank/DBJ databases.
RN [6]
RP SEQUENCE FROM N.A.
RX FLYBASE;
RL Submitted (MAR-2004) to the EMBL/GenBank/DBJ databases.
RN [7]
RP SEQUENCE FROM N.A.
RX Cadieu, Dreano, Lelaure, Mottier, Galibert F.;
RL Submitted (JUL-1998) to the EMBL/GenBank/DBJ databases.
RN [8]
RP SEQUENCE FROM N.A.
RX Benos P.;
RL Submitted (APR-1999) to the EMBL/GenBank/DBJ databases.
RN EMBL; A8003421; AAF45644.1;
RN EMBL; AL031028; CAA19845.2;
RN FLYBase; Fggn0025390; CGI4796.
DR GO; GO:0005576; C:extracellular; IEA.
DR GO; GO:0008061; F:chitin binding; IEA.
DR GO; GO:0006030; P:chitin metabolism; IEA.
DR InterPro; IPR002557; Chitin_bind_PerA.
DR Pfam; PF01607; CBM 14; 2.
DR SMART; SM00494; CHIBD2; 3.
DR PROSITE; PS50940; CHIT_BIND_II; 2.
SQ SEQUENCE 1795 AA; 194464 MW; 07F10C129BD9557B CRC64;
Query Match 18.9%; Score 1103.7; DB 2; Length 1795;
Best Local Similarity 21.2%; Pred. No. 7.8e-14;
Matches 382; Conservative 164; Mismatches 444; Indels 813; Gaps 66;
QY 3 WKTLPYILLILLISVFVIQVQSODLSSCAGRC-----G 35
DB 6 WST-----LIGLLSV-----LSTATSRVQLPSAYPPVLTGYGQPPYAYPPYG 51
QY 36 EGVSR-----DATCNCDCYNC-----OHYMECC 57
DB 52 YGYQAPPVYQPYDYNVGFVSKYGGGFFVRFYDNNRCSNRYGKPHQPPQYVYCK 111

```

```

QY 58 PPKRVCTAELCK-----GRCFSEFERGR-----BCDDAQ-----CKKYDKC 96
Db 112 PD-----CVIFSKORGLESEFNASSGRVCHVPOHDPHRRPPOCQOQREGFRPHDCKVYRC 167
QY 97 -----CPDYE-----SF-----CAEE-----107
Db 168 DKNRQOWLFPACAGTIFSPVERKCLPGQOCHSTSEISDSGSIPOQNCBLKPECEAEGETF 227
QY 108 -----HSVENQESS-----SSSSSSSSTIWK-----131
Db 228 RSPDICALYTCRLQBSGYLQTRFKPCGSNSFDLERKLCRPRSEVDCDFVGPVQVPY 287
QY 132 -----TKSSK-----136
Db 288 APQYVPPYPAAPLYEEDDYDTGAREQPALKSEKLQVAAGFEKPSLNVVVLQTTILE 347
QY 137 -----NSAANRELQK-----KLKVKDN-----153
Db 348 PSTAYHKYPAYPSYPSYSHHRRKERAENLEKEGVPKRLKLSENIVIOPETPATA 407
QY 154 -----KKNRTKKPKTPKPPVVDGAGSLDNGDFKVTTDTSST 191
Db 408 ATTREPLNDINKYQYKRYTYGTDKNDVTEAPEIKSL-----KGHLSENIIVILPETITT 462
QY 192 QHNKVTSPKITTAKPINRPSLPPNSDT-----SKETSILVNKETTIVETKETT 240
Db 463 -----TTTTTKPVWLCTPISPDPTPKPSTTAVTKSTPKISSTEQHSITTTAKITTKRPT 520
QY 241 TTNKQSTQCKEKTSAKTSQLEKTSAXDLAPTSKVLAKPTKBAETTKGPAITTKPREP 300
Db 521 VTEKTSATEKPRITVVTTTQKRSTTHNTSPDTKTITRSTLSPKITTSTSTPTST 580
QY 301 TPPTPKEPASTTKEPTPTIKSAPTTKPEPATTTKSAPTTKPEPATTTKPEAPTTK 360
Db 581 TPST-----TPSTTPST-----TPPS-----TTTTSTTTIVK-----VSTRPRTISOK 621
QY 361 EPAPTTTKEPAPTTTKSAPTTKPEPATTPKKPEPATTPKPEPATTTPKEPTTTPKEPAPT 420
Db 622 -----TIT-----ASITTKTITSPK-----TTKTIDPTSTTSKLSSTTKQTTTTTHKFTAAT 670
QY 421 -----TKPEAPTTPK-----EPAPTKKPEAPTTKPEPATTTKPEAPTTTKPE 464
Db 671 TSTEKPTTTTEKTSVTSTTKTSSTESPKSTSTGKPTTTPKPESTTTTTPKSTTTTQ 730
QY 465 SPPTKPEPATTTKSAPTTTKPEPATTKSAPTTPKPEPSTTTTKPEPATTTKPEAPTTPK 524
Db 731 TTTTPLASSTSTQPPPTTTTQPTTTTTLVTPKSTTTTTEKPTISSPKPTTTTQ 790
QY 525 -----KPAPTTPKPEAP-----TPKPEPATTTKKAPTAPEAPTTTP 563
Db 791 TTSTAPNTTKVAITTKQETPTQSTSTTIFTRKTTTNNPEPTSTKPTSTTTPKPSSTTP 850
QY 564 K-EAPATTPKLTPTTPEKLAPTTPEK-----APTPEELAPTTPEPTPTTPEEP 614
Db 851 KTSTVASSTKTTISSPKPTTEKSTENPTNSVKTSTALTSSTORASTTSEPTKIT-QNI 909
QY 615 APPTKAAAPNTPKPEAPTTPKPEAPT--TPKPEPATTPKETAPTTPKGAPTTTLKEPAP 672
Db 910 TTTTPKPTLLKSTQEAATTSTQKSVTVIITTKKATESPLTILSTEENPTPKPLRTTP 969
QY 673 TTPKKPAPKELAPTTTKEPT--STTSKAPATTPKGTAPTTPK-----EPAPTTPK 721
Db 970 TTTSVTATTTTITISESTSTSTOKPKSTTPTSTRTTPKVTTVIVSTQNTPTTTTSK 1029
QY 722 -----EPAPTTPKGTAPTTLKEPAPTTPKKAPKELAPTTTKG-----PTST-----763
Db 1030 TSTVTIITPNPSPSTQPT--TTTRQTSIT-----ASTTSIGTTRITPTTNPQNS 1078
QY 764 -----TSDK-----767
Db 1079 TSSTDLATVTRPPCPDPDSTSDKNNTACTQELQVNNLELQSPKQEQFTHTRTHTALT 1138

```

```

QY 768 -----PAPTTPKETAPTTPKEPA-----PTTP--K 790
Db 1139 GSRNVLGGQVDPYDMADPSSAEASQOATAKAPTMSTLAAAHLLQKLFHILSTTPSR 1198
QY 791 KPAPTTPETPPPTTSEVS-----808
Db 1199 EHAPTORPSPQSPSSQSRGVTIAQMARHNLATSKPFIASHRLSLIOQLASTQKRSIPPK 1258
QY 809 ---TPTTTKP-----TTIHKSPDESTPELSAETPKALENSPKPGVPTTKTPAAT 857
Db 1259 TLVTHNTTKPEPSEYDYDSEYTDEN-EVLDTQTPRAMSST-----TVAAV 1307
QY 858 KPENTTTAKTKTTERDLRTTPTTTAAPKMTKETATTT-----896
Db 1308 LPAYPST-----TTTERE-----POKTSSSPSPTKATSTTQPIETTGDLEVDSGSSDYV 1359
QY 897 -----KTTESKITATTTQVT-----911
Db 1360 NDANDISGVVNSLEARNFLSLLKORLTQIERTEAKKPATSTSTTDAPKSSSSSTSP 1419
QY 912 -STTTQDTTPPKITTLKTTT-----LAPKVTTTKTIT--TTEIMNKPEE-----953
Db 1420 ASITSESTSPVSTASRLTASKHLGPEALSRCQSLTPQSAEYVDEDDDDYMEDEPVGSS 1479
QY 954 -----TAK-----956
Db 1480 DAKEKHEGTVLISEKQAAATAKRIAPPQAQPLFQAMNLATRSBENKVDLKTSDGLQ 1539
QY 957 -----PKDRAT-----NSKATTPKQKPTKAPKPTSTKKPKTM 990
Db 1540 FRRFESSOIPTGEPNPNASIVVAEALSRRLADNVEHLPGPTVPTHTSDAKAISVSQTTI 1599
QY 991 PRVRKPKT---TPTRKMTSTMP-----ELNPTSRIAEAMLO---TTRPNQTPNSKLVEV 1040
Db 1600 -RSRESQPKSSSQTPPSISTETPATQAIQNHRTDEATQSTLEIVTQTTPKSAPPATAVPV 1658
QY 1041 NPKSEDAG-----GAEGETPHMLLEPHVFMPEVTPDMDYLERVPMQGI-----IINP 1087
Db 1659 AVROEVLVSNRPMLLAARNQTVHL-----TPISSIAARAFSNPVSHANRSINP 1707
QY 1088 MLS 1090
Db 1708 LVS 1710

```

Search completed: October 13, 2004, 12:17:01
Job time : 153.845 secs

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.
OM protein - protein search, using sw model
Run on: October 13, 2004, 11:23:49 ; Search time 22,202.2 Seconds
(without alignments)
4723.689 Million cell updates/sec
Title: SEQ1-B
Perfect score: 5826
Sequence: 1 MAWKLPYLLLSVFIQ.....DMDVLPVNPQGIINPMLS 1090
Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.1
Searched: 283416 seqs, 96216763 residues
Total number of hits satisfying chosen parameters: 283416
Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries
Database : PIR 79: *
1: Piri: *
2: Pir2: *
3: Pir3: *
4: Pir4: *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | ID | Description |
|------------|--------|-------------|--------|----------|---------------------|
| 1 | 1330.8 | 22.8 | 3020 | 2 A43932 | muscin 2 precursor, |
| 2 | 1127.2 | 19.3 | 1664 | 2 T18262 | S-layer protein - |
| 3 | 1004.7 | 17.2 | 1489 | 2 T31108 | cyst germination s |
| 4 | 1002.8 | 17.2 | 1274 | 2 T16251 | hypothetical prote |
| 5 | 992.5 | 17.0 | 7962 | 2 I38346 | elastic titin - hu |
| 6 | 987.1 | 16.9 | 3370 | 2 T45025 | muscin MUC5B, trach |
| 7 | 974.5 | 16.7 | 2187 | 2 T38226 | nascen polypeptid |
| 8 | 928.3 | 15.9 | 1367 | 1 S48478 | glucan 1,4-alpha-g |
| 9 | 901.5 | 15.5 | 3507 | 2 T34513 | hypothetical prote |
| 10 | 888.9 | 15.3 | 1188 | 2 S49315 | extensin-like prot |
| 11 | 869.3 | 14.9 | 1229 | 2 T25697 | hypothetical prote |
| 12 | 862.5 | 14.8 | 6642 | 2 T29757 | protein UNC-89 - C |
| 13 | 839.5 | 14.4 | 2897 | 2 B48666 | cell proliferation |
| 14 | 839.5 | 14.4 | 3256 | 2 A48666 | cell proliferation |
| 15 | 833.4 | 14.3 | 5762 | 2 A41819 | proline-rich pepti |
| 16 | 822.8 | 14.1 | 1151 | 2 T18535 | high molecular mas |
| 17 | 818.7 | 14.1 | 4135 | 2 T42629 | tenascin - bovin |
| 18 | 814.1 | 14.0 | 1344 | 1 A35175 | muscin 1 precursor, |
| 19 | 788.2 | 13.5 | 2232 | 2 T34434 | hypothetical prote |
| 20 | 767 | 13.2 | 5862 | 2 T03454 | ALR protein - huma |
| 21 | 764.4 | 13.1 | 1832 | 2 T31113 | muscin-like Glycopr |
| 22 | 761.8 | 13.1 | 2142 | 2 B35098 | MHC class III hist |
| 23 | 752.5 | 12.9 | 4006 | 2 T03070 | probable tenascin |
| 24 | 746.2 | 12.8 | 3942 | 2 T42730 | Basoon protein - |
| 25 | 745 | 12.8 | 2225 | 2 T56063 | hypothetical prote |
| 26 | 744.5 | 12.8 | 990 | 2 I51618 | nucleolar phosphop |
| 27 | 738 | 12.7 | 761 | 2 C84672 | hypothetical prote |
| 28 | 737.7 | 12.7 | 4548 | 1 S06657 | apoptosis(a) (EC |
| 29 | 735.1 | 12.6 | 1870 | 2 S37671 | MHC class III hist |

| | | | | | |
|----|-------|------|------|----------|---------------------|
| 30 | 734.7 | 12.6 | 1872 | 2 S36152 | MHC class III hist |
| 31 | 733.5 | 12.6 | 924 | 2 S27923 | gene IF3 protein - |
| 32 | 731.3 | 12.6 | 971 | 2 T19431 | hypothetical prote |
| 33 | 730.4 | 12.5 | 1777 | 2 T34369 | hypothetical prote |
| 34 | 730.2 | 12.5 | 2774 | 2 A43359 | microtubule-associ |
| 35 | 726.1 | 12.5 | 3938 | 2 T42761 | Basoon protein - |
| 36 | 716.2 | 12.3 | 801 | 2 T29018 | hypothetical prote |
| 37 | 715.2 | 12.3 | 1611 | 2 T38236 | hypothetical prote |
| 38 | 713.7 | 12.3 | 1630 | 2 A53577 | ascites siialoglyco |
| 39 | 702.5 | 12.1 | 3381 | 2 T42389 | versican precursor |
| 40 | 696.2 | 11.9 | 2938 | 2 T30249 | cell proliferation |
| 41 | 696.1 | 11.9 | 4957 | 2 T03455 | ALR protein - huma |
| 42 | 696 | 11.9 | 839 | 2 T75518 | hypothetical prote |
| 43 | 694.1 | 11.9 | 5105 | 2 T32650 | hypothetical prote |
| 44 | 693.5 | 11.9 | 3566 | 1 A40701 | tenascin-X precurs |
| 45 | 691.1 | 11.9 | 5170 | 2 T15348 | hypothetical prote |

ALIGNMENTS

RESULT 1
A43932
muscin 2 precursor, intestinal - human (fragments)
N:Alternate names: muscin SMUC-41
C:Species: Homo sapiens (man)
C:Date: 10-Mar-1993 #sequence revision 12-Apr-1996 #text change 09-Jul-2004
C:Accession: A49963; A45106; B45106; A43932; B3532; A61257; PQ0328; PQ0329
R:Gum Jr., J.R.; Hicks, J.W.; Toribara, N.W.; Siddiki, B.; Kim, Y.S.
J. Biol. Chem. 269, 2440-2446, 1994
A:Title: Molecular cloning of human intestinal mucin (MUC2) cDNA. Identification of the
A:Reference number: A49963; MUID:94132002; PMID:8300571
A:Accession: A49963
A:Molecule type: mRNA
A:Residues: 1-639 <GU1>
A:Cross-references: UNIPROT:Q02817; UNIPROT:Q14884; GB:L21998
R:Gum Jr., J.R.; Hicks, J.W.; Toribara, N.W.; Rothe, E.M.; Lagace, R.E.; Kim, Y.S.
J. Biol. Chem. 267, 21375-21383, 1992
A:Title: The human MUC2 intestinal mucin has cysteine-rich subdomains located both upstr
A:Reference number: A45106; MUID:93016075; PMID:1400449
A:Accession: A45106
A:Status: not compared with conceptual translation
A:Molecule type: mRNA
A:Residues: 626-1895 <GU2>
A:Cross-references: GB:M94131; NID:G186395; PIDN:AAA59163.1; PID:G186396
A:Note: sequence extracted from NCBI backbone (NCBIP:116706)
A:Accession: B45106
A:Status: not compared with conceptual translation
A:Molecule type: mRNA
A:Residues: 2037-3020 <GU3>
A:Cross-references: GB:M94132; NID:G186397; PIDN:AAA59164.1; PID:G186398
A:Experimental source: colon
A:Note: sequence extracted from NCBI backbone (NCBIP:116698)
R:Toribara, N.W.; Gum Jr., J.R.; Culhane, P.J.; Lagace, R.E.; Hicks, J.W.; Petersen, G.M.
J. Clin. Invest. 88, 1005-1013, 1991
A:Title: MUC-2 human small intestinal mucin gene structure. Repeated arrays and polymorp
A:Reference number: A43932; MUID:91358717; PMID:1885763
A:Accession: A43932
A:Molecule type: DNA
A:Residues: 1343-1350, 'L', 1352-1411, 'S', 1413-1448, 'P', 1450-1503, 'T', 1505-1915 <TOR>
A:Cross-references: GB:M74027; NID:G188863; PIDN:AAA59875.1; PID:G188864
A:Note: sequence inconsistent with the nucleotide translation
A:Note: sequence extracted from NCBI backbone (NCBIN:55749, NCBIP:55750)
R:Gum, J.R.; Byrd, J.C.; Hicks, J.W.; Toribara, N.W.; Lamport, D.T.A.; Kim, Y.S.
J. Biol. Chem. 264, 6480-6487, 1989
A:Title: Molecular cloning of human intestinal mucin cDNAs. Sequence analysis and eviden
A:Reference number: A33532; MUID:89197956; PMID:2703501
A:Accession: B33532
A:Molecule type: mRNA
A:Residues: 1816-2193 <GU4>
A:Cross-references: GB:M22405; NID:G188873; PIDN:AAA36334.1; PID:G188874
A:Experimental source: intestine
R:Jany, B.H.; Gallup, M.W.; Yan, P.S.; Gum, J.R.; Kim, Y.S.; Basbaum, C.B.

J. Clin. Invest. 87, 77-82, 1991
 A:Title: Human bronchus and intestine express the same mucin gene.
 A:Reference number: A61257; PMID:91086481; PMID:1985113
 A:Accession: A61257
 A:Status: not compared with conceptual translation
 A:Molecule type: mRNA
 A:Residues: 'T', 1925-1948, 'TTS', 1952-1954 <JAN>
 A:Experimental source: bronchus
 R.Xu, G.; Huan, L.; Khatri, I.; Sajjan, U.S.; McCool, D.; Jones, C.; Forstner, Biochem. Biophys. Res. Commun. 183, 821-828, 1992
 A:Title: Human intestinal mucin-like protein (MLP) is homologous with rat MLP in the C-b
 A:Reference number: PQ0328; PMID:92198477; PMID:1550588
 A:Accession: PQ0328
 A:Molecule type: mRNA
 A:Residues: 2328-2468 <XUG>
 A:Cross-references: GB:M86523
 A:Experimental source: small intestine
 A:Accession: PQ0329
 A:Molecule type: protein
 A:Residues: 2328-2342, 'K', 2344-2354 <XUG1>
 C:Genetics:
 A:Gene: GDB:MUC2
 A:Cross-references: GDB:120203; OMIM:158370
 A:Map position: lip15.5-lip15.5
 C:Superfamily: von Willebrand factor; von Willebrand factor type A repeat homology; von
 C:Keywords: glycoprotein; intestine; tandem repeat
 F:2766-2834/Domain: von Willebrand factor type C repeat homology <VMC>
 Query Match 22.8%; Score 1330.8; DB 2; Length 3020;
 Best Local Similarity 20.5%; Pred. No. 6.4e-19; Mismatches 495; Indels 992; Gaps 63;
 Matches 411; Conservative 108; Gaps 63;
 QY 6 LPYLLLSVFIQOVSSQDLSSCAGRCG-----EG-----YSR 40
 DB 502 LQVQLAPVQLFVTLDAQSQ--GVQGLCGNFGLEGDDFKTAGSLVEATGAGFANTWKA 559
 QY 41 DATCN-----CDY---48
 DB 560 QSTCHKDLWLDPPSLNIESANYAEHWCSSLKKTETPFGRCHSAVDPAEYKRCYDTC 619
 QY 49 NQOHVMEC-----CPDFKRVCTA-----ELSK 71
 DB 620 NQNNEDCLCAALSSYARACTAGVWLWGWREHVCNKGVDGSCPNQVFLYNLTTCQCTCR 679
 QY 72 -----GSCFESF-----ERGR-----ECDC-----DAOCKYDK 95
 DB 680 SLSEADSHCLEGFAPVDGCGCPDHTFLDEKRCVFLAKCSYHRGLYLEAGDVVVRQER 739
 QY 96 C-----96
 DB 740 CVCRDGRHLRQIRLIGOSCTAPKIHMDCSNLTALATSKPRALSCQTLAAGYHTECVSG 799
 QY 97 ---CPD-----99
 DB 800 CVCPDGLMDGRGCGWEKECFVENNDLYSSGAKIKVDCNCTCKEGRWVCTQAVCHGT 859
 QY 100 -----YESFCABEHS-----VSENOESSSSSSSSSS 126
 DB 860 CSIVSGHYITFDGKYDFDGHGCSYVAVQDYCGNSLGSLSIITENVPCGTGVTCSKA 919
 QY 127 STIWKIKSSKNSAANRELQKLKVDNKK-----155
 DB 920 IKIFMGR-----ELKLEDKRVVIQRDEGHVAYTTREVGQYLWVESSTGII 967
 QY 156 ---NRT---KKPTKPPVVDAGS---GLDNGDFV-----TT 185
 DB 968 VWDKRTIVFKLAPSYKGTGCGNFDHRSNNDFTRDHMVVSSSELDFGNSWKEAPTC 1027
 QY 186 PDTSTTQ-----HNKVTSP-----200
 DB 1028 PDVSTNPSPCLNPHRRSMAEKQCSILKSSVFSICHSKVDPKPPYACVHSDSCSDTGGD 1087
 QY 201 -----KITTAKP 207

DB 1088 CECFCSAVASYAQECTKEGACVFWRTDLCPIFCDYNNPHECEWHYEPGCRNRSFETCET 1147
 QY 208 IN-----PPPSLP-----215
 DB 1148 INGIHSNISVSVLEGCPYPCPKDRPIYEDLKKCVTADKCGCVEDTHYPGASVPTBET 1207
 QY 216 -----PNSDTSKE-----TSLTVNKET---232
 DB 1208 CKSCVCTNSQVCRPEBEGKILNQDGAFCVWEICGPNGTVEKHFNCSITRSTLT 1267
 QY 233 -----TVETRETTTNKQSTDTGKEKTSK-----258
 DB 1268 FTTITLPTTPTSTTTTTTTTTSSVLSLTTPKCLCLMSDWINEDHPSSGSDGDPED 1327
 QY 259 -----ETQSIKTSKD-----270
 DB 1328 GVGAPEDIECRSVKDPHLSLEHQHGVQCDVSVGFICKNEQDFGNGFGLCYDYKIRVN 1387
 QY 271 -----LAPTSKVLAKPTKRAETTTKGPALTTPKEPTTPKEPASTTPKEPTTIK 322
 DB 1388 CCWPMDCITTPSPPTTTPSPPTTTLPTTTPSPPTTTPPTTTPPTTTPPTTTPPTT 1447
 QY 323 SAP-TTPKEPAPTTTKSAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTT 380
 DB 1448 PLPTTTPSPPTTTPPTTTPSP-PITTPSPPTTTPSPPTTTPPTTTPSPPTTTPSPPTT 1505
 QY 381 TP-----KBPAPTTKPKAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTT 432
 DB 1506 TITTPASATTLPTTTPSPPTTTPPTTTPSPPTTTPPTTTPPTTTPPTTTPPTTTPPTT 1565
 QY 433 APTAPKAPPTTTPKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTT 492
 DB 1566 PTTTTTPPTTTPSPPTTTPSPPTTTPPTTTPSPPTTTPPTTTPPTTTPPTTTPPTTTPSP 1624
 QY 493 KSAPTTTPKEPPTTTPKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTT 551
 DB 1625 ---PITP---PTSTTLPTTTPSPPTTTPPTTTPSPPTTTPPTTTPPTTTPPTTTPPTTTPPTT 1677
 QY 552 PTAPKEPAPTTTPKEAPTTTTPKEAPTTTTPKEAPTTTTPKEAPTTTTPKEAPTTTTPKEAPTT 611
 DB 1678 TITTPSPITTP--SPPTT---TWTPS--PTTPSPPTTTPPTTTPPTTTPPTTTPPTTTPPTT 1728
 QY 612 EBPAPTTTAAAAPNTKBP-----PTTPKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTT 659
 DB 1729 TTPSPPT-----TPSPPTTMTLPTTTPSSPLTTPPTTTPPTTTPPTTTPPTTTPPTTTPPTT 1781
 QY 660 -----KG-----661
 DB 1782 CVPLCNWTCGLSDGKFNHFKPGDTELICDVCPCGMAANISCRATWYDPVPIGQLGQIVV 1841
 QY 662 -----TAPTTLKEPAPTTTPKPKAPKEL 683
 DB 1842 CDVSVGLICKNBQKPGGVIPIMAFCLNIEINVQCECVTQPTM---TTTTTENPTPTPI 1898
 QY 684 APTT--TKPETSTTSDKP---APT-TPKTAFTTPKEPAPTTTTPKEPAPTTTTPKEPAPTTT 736
 DB 1899 TTTTTPPTTTPSTQSPNGLQAPPTTISTITVPTPTPTGCTGPT-TTPTITTTTPT 1957
 QY 737 KEAPPTTPKPKAPKELAPTTTTPKPTST-TSDKAPTTTKEPAPTTTKEPAPTTTTPKPP---792
 DB 1958 PPTPTGTGTPTVTLITTTTTMTPTPTSTKSTVPTTITTTTITVATPTGTGTPTMI 2017
 QY 793 -----APTTPETPTTSEVSTPTTK-----EPTTIHKSDESETPELSAETPRALENSP 843
 DB 2018 PSTTTTTVPTPTTIGTGPTTHTSTADIAELITSNPPPESSIPQTSRSTSSPLTSTT 2077
 QY 844 KBPGVPTTKTAAKPEMTTAKOKTTERDLRTPETTTAAAPKMTKETATTTTEKTESKI 903
 DB 2078 LLSTLPPAEMTSTAPPSTAPTITTSCHTLSPSPSTTTPPGTPTRGTTT--GSSSAP 2135
 QY 904 TATTITQVSTTQDTPPKIT--LKTTLAPKVTTKTITTTTEINMKPEE-----953

Db 2136 TPSTVQTTTSAWTPPTPLSTPSIIRTTGLRPYPSSVLICVNDTYVAPGEEVNGTY 2195
Qy 954 -----TAPKDRATNSKATTPKQPKPTKAPKPKSTKPKTMP 991
Db 2196 GTCYFVNCGLSCTLEFYNWSCPTSPPTPSK-STPTSPKSPSTPKPTGTRKPPCP 2254
Qy 992 RVRKPKXTTTPRKMTSTMPELNPTSRIAEAMLTQTTTRNQNTNSKLVEVP----- 1042
Db 2255 DFDPPP-----QENETWLCDFM-AICKYNNVTVEIVKVECEPPMPPTCSNG 2300
Qy 1043 -----KSEDAGGA-----EGETHMLLRPHVFMVETTP 1070
Db 2301 LOPVRVEDPGCWHWCEDCYCTGWDGPHYVTFDGLYYSQGNCTYVLVE-----EISP 2354
Qy 1071 DMD-----1073
Db 2355 SVDNFGVYDNYHCDENDKVCERTLIVRHETQEVLLIKVHMMPQVQVNRQVAALPY 2414
Qy 1074 -----YIPRVNQIILN 1086
Db 2415 KRYGLEVYQSGINYVVDIPELGLVLS 2440

RESULT 2
T18262
S-layer protein - Clostridium thermocellum
C:Species: Clostridium thermocellum
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 09-Jul-2004
C:Accession: T18262
R:Fujino, T.; Beguin, P.; Aubert, J.P.
J. Bacteriol. 175, 1891-1899, 1993
A:Title: Organization of a Clostridium thermocellum gene cluster encoding the cellulose
e.
A:Reference number: Z18847; MUID:93209931; PMID:8458832
A:Accession: T18262
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-1664 <FUI>
A:Cross-references: UNIPROT:Q06852; EMBL:X67506; NID:G296879; PID:G296881; PIDN:CAA47841

Query Match 19.3%; Score 1127.2; DB 2; Length 1664;
Best Local Similarity 21.4%; Pred. No. 2,9e-15;
Matches 369; Conservative 127; Mismatches 427; Indels 799; Gaps 68;

Qy 4 KTLPIYLLLLSVFIQVSSDLSACRGCYGRDTCNCDYNCQHYMECCP----- 58
Db 6 KVLISILLTLL-----IISTSYNMS-----FAEATPSIEMV 37
Qy 59 -----DFKRVCTAELSK-----GRCFESFERGECDC 86
Db 38 LDKTEVHVGDVITAIKVNIRKLAGYQLNIKFDPEVLQVDPATGEEFT----- 87
Qy 87 DAQCKYDKCCP-----DYSFCAEEH-- 108
Db 88 -----DKMPVNRVLLTNSKYGTPVAGNDIKSGIINFATGYNNLTAYKSSGIDEHTG 140
Qy 109 -----SVSENOBSSSSSSSSSSSTIW-----KIKSS 135
Db 141 IIGETGFKVLKQNTSIRFEDTLSPGGAISGSLFDWDAEITGVEYIQDILIVVEAPL 200
Qy 136 KNSAANRELQK-KLKVD-----NKN-----RTKKKPT----- 163
Db 201 KDAVALELDKTKVKVGDIIITATIKENNMKNFAGYQLNIKVDPTMLBALELTGSAIAKR 260
Qy 164 -----PKPPVDEAGSLDN----- 178
Db 261 TWPVTGGTVLQSDNYGKTTAVANDVAGGIINFAPAEAYNSLTKYRETVGVAEETGIIGKIGFR 320
Qy 179 -----GDFKVTTPDT-----STQHNKV-----STSPK 201
Db 321 VLKASTAIRFEDTTAMPAGIETGYWFDYNGENIKGYVWQGEIVAEGBEPGEPTSP 380
Qy 202 ITTAKPINRPELP-----PNS-----DTSK----- 222

Db 381 VPTEPVDPTPTVEEVPSELDDSYVIMELDKTKVKVGDIIITATIKENNMKNFAGYQLN 440
Qy 223 -----ETSUTVNKETTVEKETTNTKQTSDDGKEKTS----- 256
Db 441 IKYDPTMLEALELTGSAIAKRTWPVTGGTV-----LQSDNYGKTTAVANDVAGIINFA 495
Qy 257 -----AKETQSIEKTSADLAPTSKVLAKPTPKAETTT----- 289
Db 496 EAYSNLTKYRETVGVAEETGIIGKIPVLKAGSTAI-----RFEDTTAMPAGIETGYMFD 550
Qy 290 -----KGPALTTP-----KEPTPT-TPKEPASTTPKEPTT----- 319
Db 551 WYGENIKGYVWQGEIVAEGBEPTEBPVTEPVDPTPTVTEEPVSELPSVIMELD 610
Qy 320 ----- 319
Db 611 KTKVKEGDVIAIRVNNIKNLACYQIGIKYDKVLEAFNIEGTGDI-DEGTWPAVGTL 670
Qy 320 ----- 329
Db 671 KNRDYLPTGVAINNVSKGILNFAAYVYVFDYREEGKSEDTGIIGNIGFRVLKAEDETIR 730
Qy 330 -----EPAPTTKTSAPTTKKEPAPTTPKKEPAPTTP 359
Db 731 FEELESMPGSDIGTYMLDYNLRISGYVYIQAPAKAAS-----DEPIPTDTPSDEPTPS 785
Qy 360 KEAPAP-----TTTKEPAPTTPKKEPAPTTPKKEPAPTTPKKEPAPT-----TP 406
Db 786 DEPTPSDEPTPSDEPTPSDEPTPSDEPTPSDEPTPSDEPTPSDEPTPSDEPTPSDEPT 845
Qy 407 K-EPTPT-TPKKEPAPT-TPKSAAPTTPKKEPAPTAPKKA-----PTTPKKEPAPT-TPKKEPAPT 460
Db 846 SDEPTPSDEPTPSDEPTPSDEPTPSDEPTPSDEPTPSDEPTPSDEPTPSDEPTPSDEPT 905
Qy 461 TKEPSPTTPKKEPAPTTPKSAAPTTPKKEPAPTTPKSAAPTTPKKEPAPTTPKKEPAPT 520
Db 906 TPSEPTPSDEPTPS-----DEPTPSDEPTPSDEPTPSDEPTPSDEPTPSDEPTPSDEPT 962
Qy 521 TTPKKA-----PTTPKKEPAPT-TPKKEPAPTTPKKEPAPTAPKKAAPTTPKE 565
Db 963 SDEPTPSDEPTPSDEPTPSDEPTPSDEPTPSDEPTPSDEPTPSDEPTPSDEPTPSDEPT 1019
Qy 566 TAPTPPKLTPTPKELAPTPPKAPAPT-----TPEELAPTPPEEPTPT-TPKEPAPTTPKA 621
Db 1020 -----EPTPSDEPTPSDEPTPSDEPTPSDEPTPSDEPTPSDEPTPSDEPTPSDEPT 1061
Qy 622 AAPNTPKEPAPTTPKKA-----PTTPKKEPAPT-TPKKEPAPTTPKKEPAPTTPKKEPAPT 678
Db 1062 DEPTPSDEPTPSDEPTPSDEPTPSDEPTPSDEPTPSDEPTPSDEPTPSDEPTPSDEPT 1119
Qy 679 APKELAPTTTKEPT-STTSCKAPAPTTPKAPAPTTPKKEPAPTTPKKEPAPTTPKKEPAPT 737
Db 1120 TPSP-EPTPSDEPTPSDEPTPSDEPTPSDEPTPSDEPTPSDEPTPSDEPTPSDEPTPS 1172
Qy 738 EPAPT-TPKKAAPKELAPTTTKEPTSTTSCKAPAPT-----TPKKEPAPTTPKKEPAPT 792
Db 1173 EPTPSDEPTPS-----PTDTPSDEPTPSDEPTPSDEPTPSDEPTPSDEPTPSDEPTPSDEPT 1226
Qy 793 APPTPEPTPTTSEVSTPTTKEPTTHKSPDESTPELSAEPKALENSKPEPGVPTTK 852
Db 1227 IPTDTPSDEPTPSD-----EPTPSDEPT-----PSDEPTPSDEPTPSDEPTPSDEPT 1274
Qy 853 TPAATKPEMTTAKDKTTERDLATTPETTTAAAPKMTKETATTTKTESKITATTTQVTS 912
Db 1275 PSDEPTPSDEPTPSDEPTPSDEPTPSDEPTPSDEPTPSDEPTPSDEPTPSDEPTPSDEPT 1334
Qy 913 TTTQDPTTPFKITITLTKTTLAPKVTITTKTITTTIMNKPKEATKPKDRATNSKATPKPKQ 972
Db 1335 TPSEPTPSDEPT-----PSDEPTPSDEPTPSDEPTPSDEPTPSDEPTPSDEPTPSDEPT 1369
Qy 973 -KPTKAPKPTSTYKPKTMRVKPKTPTTPTPKWTS-----T 1008

Db 1370 PTPSTPREPT-----PTTPTPTPTTSGSGSGGGGGGGGTVPT 1416
Qy 1009 MPELNPTSRIAEAMLTTRRNQTPNSKLVBNPKSBDAGAGETPHMLLRPH---VFM 1065
Db 1417 SPTPTPTS-----KTSPTAPTEIE-EPTPSDVPGAIGCEHRAVLRGYPDGSFR 1464
Qy 1066 PE-----1067
Db 1465 PERNITRAEAAVIFAKLGADESQAQSPSYDLATHWAAWAIKATSOGLPKGYPDG 1524
Qy 1068 -VTPD-----MDYLPRVFNQI-----IINP 1087
Db 1525 TFKPDQNTAEFAIVLHFLTKVKGQIMSKLATIDISNP 1565

RESULT 3
T31108
cyst Germination specific acidic repeat protein precursor - Phytophthora infestans
C:Species: Phytophthora infestans (potato late blight agent)
C:Date: 22-Oct-1999 #sequence_revision 22-Oct-1999 #text_change 09-Jul-2004
C:Accession: T31108
R:Goernhardt, B.
Submitted to the EMBL Data Library, April 1998
A:Reference number: Z20986
A:Accession: T31108
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-1489 <GO>
A:Cross-references: UNIPROT:O96449; EMBL:AF061185; NID:G3851513; PID:G3851514; PIDN:AACT
C:Genetics:
A:Gene: car90

Query Match 17.2%; Score 1004.7; DB 2; Length 1489;
Best Local Similarity 26.2%; Pred. No. 6.5e-13;
Matches 406; Conservative 60; Mismatches 422; Indels 663; Gaps 77;
Qy 41 DATCNCDNCOHYMEC-----CPDPKRVCTAELSKGRCFSEFGRGDCD--- 86
Db 32 DATYSLSHSCG--IPCSGVGAEPVGTACPKAGDVATS-----DCQPY 71
Qy 87 -----DAQCK-----KYDKCCPDYESFCAEHSVSENQSSSSSSSS 124
Db 72 LLSYNGAVCVAPVDAECALIHDDMWGCEFPK--TGYSVAEAEIAVNGESSG----- 123
Qy 125 SSSTIWKIKSSKNSAANRELOKLVKDKNKN---RTKKKPTPKPV----- 168
Db 124 ----W-----GTGHDEV--VQVGDEBEIPARVNYDTVDTPIGVNCEVATATQ 168
Qy 169 -----VDEAGSGLONGDF-----KV----- 183
Db 169 GHATEGKYDTPSTGTGTGDNTHYGSTTTEGVTKGGYGPDTAKVIDGETVLDVPTG 228
Qy 184 -----TTPDTSTT-----QHNKVSTSPKI--- 202
Db 229 ITEILEDGTPPGYCTTDDGGTTTGGYTTVDNTHETEGAGGYDAGTREEVSTPTVGYS 288
Qy 203 -----TTAKSIN-----PRSLP----- 215
Db 289 TEETEGQVHTGGYBPSDEAPTEGTYVPRBEETAPSEDITYAPREVTVYAPTEKPYD 348
Qy 216 -----PNSDT----- 220
Db 349 VEETTVTEETSYATKSETNAPTEPMHYAIEKPCDTEVTMYAPTEETTVYAPTEETTYA 408
Qy 221 -SKETSLTVNKETTYE--TKETT--TNKOTSDGKKTTSKAKETOSIEKTSKADLAPSKV 277
Db 409 PTEETTVAPTEETPYEPTTEETTYTTEETTVAPTEETT-----YAPTEKT 453
Qy 278 LAKPTPKAETTTKGPALETPPKETPTT-----PKE--- 307
Db 454 TVAPT-----EETTVAPTEETPYEPTTEETTYTTEETTVAPTEETTYASTTEETTVAPTEET 510
Qy 308 --PASTTPKPTPTTIKSAPTTPKBPAPT--TTKSAPTTPKBPAPT--TTKEPAPTTPKBP 362

Db 511 YAPAEETPYEPTTEET--TVAPTEETTVYAPTEETTVAPTEETTVYAPAEETPYEP 569
Qy 363 APTTKKEPAPTTTKSAPT-----TPKEPAPTTPKKAPAPTTPKBPAPTTPKEPTPT 412
Db 570 TEETTVAPTEETT--YAPTEETMYAPIETETTVAPTEETTVYAPAEETPYEPTTEETTVAPTEE 628
Qy 413 TKKEP-----AP-----TKKEPAPTTPKEP-----APTAKKAPATPKKPA 449
Db 629 TTVAPTEETTYASTBETTVYAPTEETTVYAPAEETPYEPTTEETTVYAPTEETTVYAPTEETTYA 688
Qy 450 PTTPKBPAPT--TTKEPSPPTPKP-----APT--TTKSAPT-----TKKEP--- 487
Db 689 PTEETTVAPTEETTVYAPAEETPYEPTTEETTVAPTEETTVYAPTEETMYAPIETETTVYPTTEE 748
Qy 488 ---APT--TTKSAPT--TKKEPSPPTTKP-----APTTPKEPAPTTPKKAPATPKKPA 535
Db 749 TTVAPTEATYAPTETPYEPTTEETTVYAPTEETTVYAPTEETTVYAPTEETTVYAPTEETTYA 808
Qy 536 PT--TPKEPAPTTPKKAPATPKP-----APTTPKSTAPTTPKKLTPTTP 579
Db 809 PTEETPYEPTTEETTVYAPTEETPYEPTTEETTVYAPTEETTVYAPTEETTVYAPTEETTVAPTEE 868
Qy 580 EKLAPT--TPKEP-----APTTPRELAPT-----TPKE 605
Db 869 TTVAPTEETPYEPTTEETTVYAPTEETTVYAPTEETTVYAPTEETTVYAPTEETTVYAPAEETPYE 928
Qy 606 PTTTTPKEB-----APTTPKAAANTPKBPAPT--TPKEPAPTTPKBP-----APTTP 651
Db 929 PTEETTVYAPTEETTVYAPTEETTVYAPTEETTVYAPTEETTVYAPAEETPYEPTTEETTVAPTEE 988
Qy 652 KETAPTTPKGAP-----TTLKBPAPTTPKBP-----APKE---LAPT----- 686
Db 989 TTVAPTEETMYAPIETETTVYAPTEETTVYAPAEETPYEPTTEETTVYAPTEETTVYAPTEETTYA 1048
Qy 687 ----TTKEPTSTTSOKPA-----PTTPKGAPTTPKBPAPTTPKBPAPTTPKGPAPT-- 734
Db 1049 STEETTVAPTEETTVYAPAEETPYEPTTEETTVYAPTEETTVYAPTEETTVYAPTEETTVAPTEE 1108
Qy 735 TLKEPAPTTPKBP-----APKE---LAPT-----TTKGPTSTTSOKPAPTTPKE 775
Db 1109 TTVYAPAEETPYEPTTEETTVYAPTEETTVYAPTEETTVYAPTEETTVYAPTEETTVYAPTEAT 1166
Qy 776 TAPT--TPKEPAPTTPKBP-----APTTPPTPTTSEVSTPT--TTKEPTTIHKSDE 825
Db 1167 YAPTEETTVAPTEETTVYPTGTEETTVYAPTEETTVYAPTEETTVYAPTEETTVYAPTEETTVAPTE 1224
Qy 826 STPELSAETPKALENSPKPEGPVTTKTPAATKPEMTTAKDKTTERDLRTPTTPTTAAP 885
Db 1225 PTEETTVAPT-----BETTVYPTTEETTVAP-----TEETTVAPTEETTVAP 1265
Qy 886 KMTKET-----ATT--TEKTESKITATTTQVTSST--TQDST--PPKAIT--LK 927
Db 1266 --TEETMYAPIETETTVYPTGTEETTVYAPTEATYAPTEETTVYAPTEETTVYAPTEETTVYAPTE 1323
Qy 928 TTTLAPKVTT-----KKTITTEIMNKPEETAKPKRATNSKATTPKQP 972
Db 1324 EITVAPTEETTVAPMEETPYEPAESTSVSTKEPCNTEETFTDEPTDEPTDEPSDEPTDE 1383
Qy 973 ---KPTKAPKPTSTTKKPTMPRVKRP-----PCDNOGINGIGVENKVRNNAGIYNTTPGPRNSQ 1432
Db 1384 PTDEPTDLP-----TDEPST-----MTSTMPELNPTSRIAEAMLTTPTRNQTNP 1034
Qy 1005 -----SDSVCELTTSTSDREDOQN 1476
Db 1433 SWHSCCRSCYNDPICHAFSHQTS-----SDSVCELTTSTSDREDOQN 1476

RESULT 4

T16251
hypothetical protein F35A5.1 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 20-Sep-1999 #sequence_revision 20-Sep-1999 #text_change 09-Jul-2004

C:Accession: T16251
 R:Reimbach, D.
 A:Description: The sequence of C. elegans coemid F35A5.
 A:Reference number: Z16485
 A:Accession: T16251
 A>Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 1-1274 <LEI>
 A:Cross-references: UNIPROT:Q20007; EMBL:U46675; NID:g1166613; PID:g1166621; PIDN:AAB526
 A:Experimental source: strain Bristol N2; clone F35A5
 C:Genetics:
 A:Gene: CESP:F35A5.1
 A:Map position: X
 A:Introns: 1272/2

Query Match 17.2%; Score 1002.8; DB 2; Length 1274;
 Best Local Similarity 23.1%; Pred. No. 5.7e-13;
 Matches 330; Conservative 99; Mismatches 309; Indels 692; Gaps 75;

QY 158 TKKKTTP-----KPP-VVDEAGSLGNDGFKVTPDSTTOHNVKSTSPKITTAKP 207
 DB 2 SEAPTPIKNPAKKWKPWESVDE----- 25

QY 208 INPRSLPNSDTSKETSITVNKETTETTTNNKOTSDGKEKTTSAKETQSIKTS 267
 DB 26 -----EEWEVDEETAPSK-----LEKPSLR----- 49

QY 268 AKDLAPTQKVLAKTPKA-----ETTKGPAITTP-----KEPTP- 302
 DB 50 -KD-APTKPV-----PSGAPSPVPIKNPVKKWKPAPWEDDEPBEAAPVPAKKVDPSPK 104

QY 303 TTPKSPASTTKE-----PTPT-----TIKSAPTTP 328
 DB 105 KVPKPRDASPKKIWAKEPBTLPVPTPVKNVKKFAPWEDDEVDVDDKADPTVP 164

QY 329 -----KEPA-----PTTKSAPTTPKEP-----APTTPKEPA-----PTT 358
 DB 165 AKKTPVLKKKEPAAAKRDPSPKAAAPSKBHDPIVPTPIKNPAKKWKPWEDDEVTE 224

QY 359 P-KEPAPTTPKEPA-----PTTKSAPT----- 380
 DB 225 EIKPEPATRKVPALKKKEPSTSVKPSVSDPSPTKKVPVKKEPVEVPTPIKNPTKKWKPW 284

QY 381 ---TP-----KEPAP-----TTPKPAPTTPKEPAPTTPK 407
 DB 285 EDETVEEVKEPVPVEKAPVKKDPAKAPAKARDSPSKAPKVEPSSPVP-PTPVK 343

QY 408 EPT-----PTTKGSPAPTTP-----EPAPTTP-KEPAP----- 434
 DB 344 NPVKKYKPPWEVDDEPABEVKPSAPEKKTPLKKEPEPSSPTTSSDPSPKAAPVKP 403

QY 435 ---TAPKAP-----TTPKAP-----TTPKEPAPTTP-- 453
 DB 404 ROSSPKKATPLOADPKAOEVPTPVKNPVKKYKPPWEVDDEPVEVQPEAPAKKTPVL 463

QY 454 ---KEPA-----PTTTKEP-----SPTTP-KEPA----- 473
 DB 464 KRKEPAAKDTAKPAISKTPTETPEKDPVKPRDSSPKKVAAPDSQAAPATPVKNPVKKWR 523

QY 474 -----PTTKSAPT--TTKEPAP-----TTTKSAPTTPKEPSPITTK 509
 DB 524 PPWEDEETPADDVSKPTDAKKTPLAKKDPAPAKESLPKADTK-APAKPRDPSP---KK 579

QY 510 PAPTTPKEPAPTTPK-----PAPTTPKBPAPTTPKEP-----APTTPKK 549
 DB 580 VAPTAPKKTPLAKKEPAGPADSKTKEPKSKPRDPSPKKAPVPAKVPKTEVAPAAVKK 639

QY 550 PAP-----TAPKEPAPTTPKETAATTP-----KKLTP-----TTPKEKLP 584
 DB 640 PEPISKPKDTPAKKAEPSNP--WSPPTPVKNPVKKWKPWEDDDAPAKPVSLPEPEKKT 697

QY 585 TTPKEPAPTTP-----ELAPTTPKEPPT--TTPKEPAPTTP----- 618

Db 698 VLAKK-APTKPDSEAAADPVSGPSSKDPKLAKKAPVKPRDPSPMKAVP-IKPAKTEVPPA 756
 QY 619 -----PK--AAADNTPEKAPTTPKEP----- 638

Db 757 VVKPBPVAKSRDPSPKKAAEENSVPVP-PTPVKNPVKKWKPWEDDDAPAPVAVPEP 815
 QY 639 -----APTTPKEPAPTTPKEPAPTTPKGTATTTIKBPAPTTP-PKKAPKE----- 682

Db 816 EKKTTPVLAKKTPVKPRDPSPKKAVPAKPSKTKTDAPPVSVKVPKPEVSKPEPSKAEFNS 875
 QY 683 --LAPTTPKEPT-----STSDKEPAPTTPKGTAPT--TPKEP-----APTTPKEP 723

Db 876 PVVPTPVKNPVKKWKPWEDDDDEPTEEVKPSPEKKTPLVAKKEPEKPKDAPKVAAP 935
 QY 724 APTTPKGTATTTIKEPAPT-----TPKKAP-----KELAPTTPTKGPTS-----TTS 765

Db 936 RDPSPKAVPE--KEPAKAAKPRDLSPKCAIPIANTQEAAPTVPKNPVKKWKPWEDD 993
 QY 766 DKP-----APTTPKET-----APTTPKEPAP-----TTPKAPAPTTPPTTSEVSTP 810

Db 994 DEPAEPVSAPEPEKKTPLVAKKAPKPRDPSPKKAAAPVAAKPKTIPEVP-----P 1044
 QY 811 TTTKEPTTIHKSP-----DESTEPESA-EP---TPKALENSPKPEGV-----P 849

Db 1045 TPVKNPVKKWKPWEDDDDEPSEPVSAPEPEKKTPLVAKKAPTATKPDSEAAADPVSGP 1104
 QY 850 TTTKTPAAT-----KPEMTTAKDKTTERDLRTTPETT--TAAPKMTKETATTTTEKTES 901

Db 1105 TSDPKLSKKAPVEKPKPTTDPKDDKLKPSPAKKPEKAPKAPKWKVPWDDDDPDEPA 1164
 QY 902 KITATTTVSTTQTDTTPFKITTLKTTLLAPKVVTTTKTITTTIMNKDEETAKPKDRA 961

Db 1165 DFTVPAPSKPDTEDPADPLG-----GPKTKDPK-----LNKKAPAEKPTK- 1206
 QY 962 TNSKATTPPKQKPTKAPKPTSTKKKPTMPRVKPK-----TTPTPKMTSTM 1009

Db 1207 -----PKEKEVSKPEPKPTTEPPKP--AAPKKWKPWEDDDDEPDEADFTMPAPKK----- 1253
 QY 1010 PELNPISRIAEAMLQITTRNQTPNSKLVENVKPSD-----AGAGEETP 1055

Db 1254 -----PTEDPADSLGGPKPKDP 1271

RESULT 5
 I38346
 elastic titin - human (fragment)
 C:Species: Homo sapiens (man)
 C:Date: 29-May-1998 #sequence_revision 29-May-1998 #text_change 09-Jul-2004
 C:Accession: I38346
 R:Labeit, S.; Kolmerer, B.
 Science 270, 293-296, 1995
 A:Title: Titins: giant proteins in charge of muscle ultrastructure and elasticity.
 A:Reference number: A57430; MUID:96026330; PMID:7569978
 A:Accession: I38346
 A>Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: mRNA
 A:Residues: 1-7962 <RES>
 A:Cross-references: UNIPROT:Q10465; EMBL:X90569; NID:g1017426; PIDN:CAA62189.1; PID:g1017426
 C:Genetics:
 A:Gene: GDB:TIN
 A:Cross-references: GDB:127867; OMIM:188840
 A:Map position: 2q31-2q31

Query Match 17.0%; Score 992.5; DB 2; Length 7962;
 Best Local Similarity 13.1%; Pred. No. 1.2e-11;
 Matches 405; Conservative 108; Mismatches 429; Indels 2145; Gaps 95;

QY 4 KTIPIYLLLLSVFVICOVSSQDLSSCAGRGSGYRDATCNC----- 46
 DB 4815 KNPFDFIRLAPDAV-----VGESADFECHVTGTQPIKVSWAKDSREI 4858

QY 47 -----DYNQHYMECCPDFKRVCTAELSCKGRCP-ES 77

4859 RSGGYQIYSLSENSAHLTVLVKVDGSGQYTCYAVNEVGKD---SCTAQLNIKERLIPPS 4915
Db
78 EERGECDDAQCKYKDYCCPDYSEFCABEHSVENQESSSSS-----SSSSSTI 129
QY
4916 FTK-----RLSETVEETEGNFKLEGRVAGSQITVA 4947
Db
130 WKIKSSKNSAANRELQKLVKDNKNRTK----- 159
QY
4948 W----YKN---NIEIQPTSCNCEITFPKNTLVLOVRKAGMNDAGLYTCVNSDAGSALCTS 5000
Db
160 ---KKPTKPPVDE-----AG----- 173
QY
5001 SIVIKER-KKPPVFDQHLTPVTVSEGEYVQLSCHVQSGEPIRIQWLKAGREIKPSDRCSF 5059
Db
174 ---SG-----LDNGDF-----KVTTPTDSTTQHNNKVSTSPKITTAKPINRPSL 214
QY
5060 SPASGTAVLELRDVAKADSGDYVCKASNVAGSDT-----TKSKVT-----IKDXPAV 5106
Db
215 PPNSD-----TSKETSJTVNKT----- 233
QY
5107 APATKKAADVGRLEFPVSEFQIRVVEKTTATFIKAVGGDPPIPNVKNWTKGWRQLNQGGRV 5166
Db
234 ---VETKETTIN-----KOTSTDG----- 250
QY
5167 FIHQGDEAKLEIRDTTKIDSLGRVAFNEHGEIESNVNLQVDERKKQEKIEGLRAML 5226
Db
251 -----KEKTT 255
QY
5227 KKTPIKXGAGBEEIDIMELLKNVDKPEYKARYGITDFRGLLOAFELLKQSEET 5286
Db
256 SAKETQSTEK-----TSKAD----- 270
QY
5287 HRLEIEIERSERDKPEELVSVFIQRLSOTEPVTLIKDIENQTVLKNDNDVAPEIDI 5346
Db
271 -----LAPTSK----- 276
QY
5347 NYPEIKLSWYGTCKLEPDSDFEISIDGDRHTLVKNCQLKQCNVRLVCGPHIASAKLT 5406
Db
277 ----- 276
QY
5407 VIEPAWHRLODVTLEKQGTCTMTVQFVSPNVKSEWFRNGRILKPOGRHKTEVEHKVHL 5466
Db
277 ----- 276
QY
5467 TIADVRAEDQGYTKYEDLETSALRLAEPIQFTKRIQNIUVSEHQSATFCEVSFDD 5526
Db
277 -----VLAKPTPKAETTTKGPA 293
QY
5527 AIVTWYKGTBELTSQKYNFRNDGRCHVTHTHNVTPDDEGYSVIARLEPRGEARSTAE 5586
Db
294 LTTPEK-----TPET----- 307
QY
5587 YLTTKELKLELKPDPIDPSRVPIITMPIRAVPPEIIPVAVPVPVLLLPTEKKPPPKR 5646
Db
308 -----PASTTPKE-----PTPTTI----- 321
QY
5647 IEVTKAVKDAKKVAKPEMTPREIEIVKPPPTTLIPAKAPEIIDVSSKAEVVKIMT 5706
Db
322 ----- 321
QY
5707 ITRKKEVQKEBAVEYKQAVHKRVFIESFPEPYDELEVEPYTEPFQPYVEEDDY 5766
Db
322 ----- 321
QY
5767 EEIKVEAKKEVHEEWEEDFESGQYBYREGEYDEGESEWEEAYQEREVIQVQKEVVEESH 5826
Db
322 -----KSAPTTPK----- 329
QY
5827 ERKVPKVPKAPPPPKVVKVIEKTSTRMBEERKQVTVKVPVSKIVPQKPSRT 5886
Db
330 ----- 329
QY

5887 PVQEEVIEVKVPAVHTKKMVISSEKMFASHTTEESVSVTVPEVQKEIVTEEKIHVAVSKR 5946
Db
330 -EPAPTTTASAPTTPKEPA-----PTTTKEPAPTTPKEPAPTTPTTKEPAPTTPKAP 379
QY
5947 VEPDP-----KVPELEKAPAEVAVPPIPKVPEPPAPKVPVP-----KKVPESKKVP 5997
Db
380 TTPKEPA-----PTTPKKPAP-----TTPEK 400
QY
5998 VPKEPEAAPPKVPPEVKPEBEKIPVPVAKKEAPPKVPPEVQGVVTEBKITITVQRE 6057
Db
401 --PAPTTPKEPTTPKE--PAPTTPKEPAPTTPKEPAPTTPKAP----- 442
QY
6058 ESPPPAVPSIPKKKVPPEERKVPPEEVEVPPPKVPA--LPKKVPPEEKVAVPVPVAKKA 6115
Db
443 ----- 442
QY
6116 PPRAEVSKTVVEEKRFAVEEKLSPAVPQVTRHEVSAEEEWSYSEEEBGSISVVR 6175
Db
443 -----TTPKEPA 449
QY
6176 BEEREEEAETVEYVMEPEEYVVEEKLHIISKRAEAPAEVTEROEKIVLKPPIA 6235
Db
450 -----PTTPKE-----PAPTTPKE--PSPTTPKEP--APTTPKAPTTP--KE 486
QY
6236 KIEEPPPAKVPKAPKIIVPEKKVPAPVPKKEKVPVPKVPPEEPKVPKVPKVPKVPK 6295
Db
487 PAP-----TTTKSAP-----TTPKEPSPTTKB-----PA 511
QY
6296 PLPAKVTEKHMOTQOEKVLVAVTKKEAPPKARVPEEPKRAVPEEKVLKLPKREEP 6355
Db
512 PTT-----PK-EPAP----- 520
QY
6356 KVTEFRKRVVKEEKVSIAPKREPOPIKEVTIMEKERAYTLSEBAVSQOEYEEYEE 6415
Db
521 -----TTPKKPAPT--TTPKEPAPTTPKE- 541
QY
6416 YDKPEEYEPTEYDQYEEVEEREYEEYITEPEKPIPVKVPPEEPVPTPKAP 6475
Db
542 -----PAPTTPKAPTAPKEP----- 558
QY
6476 PAKVLKXAVPEEKVPVPIPKKLKPPPKVPPEEPKVPPEEKIHSITKKEKQVTEPAAKV 6535
Db
559 -----APTTPKETAP----- 568
QY
6536 PMKXRVVAEEKVPVPRKEVAPPVVRVPEVKLEPEEVAPEEVEVTHVEEYLVVEEVEVI 6595
Db
569 -----TTPKKLTPPTPEKLPAPTTPKEP-----APTPELA 599
QY
6596 HBEESFITESEVVPVIVKVP-PEVPRKEVPPEEKVPVVPKKEAPPKVPPEVPKPKPEKV 6654
Db
600 PT-----TPEEPTPTTPKEPAPT-----TPKAAAPNTPKPEPAPTTPKEPAP 640
QY
6655 PVLIPEKKEPPPAKVPPEVKVPPEEKVPVVPVVKVEAPPKVPPEVKVPV-EKKVPVP 6713
Db
641 T-----TPKEPAPTTPKETAPTTPKGT--APTTLKEPAPTTPKK----- 677
QY
6714 APKVEAPPKVPPEVKKLIPEEKFTVVPKKEAPPKVPKREPPVPVVALPOEEVVL 6773
Db
678 ----- 677
QY
6774 FEEIIVPEEVLPEEVEVLPEEVEVLPEEVEVLPEEVEVLPEEVEVLPEEVEVLPEEVEVL 6833
Db
678 -----PAP-----KELAP----- 685-
QY
6834 PEEVLPVKVPVPAVPAVPEIKKVKVVPKKEBAPPKVPPEVPKKEKRIILPK 6893
Db
686 -----TTPKEP-----TSSTSDKAPTTPKG 706
QY
6894 EEEVLPEVEEPEEPEEPISEEEIPEEPPSIESEVEVAPPVPEVIEKAVPEAPTVPKVK 6953
Db
707 TAPTTP-----PKE--PAPTTPKSPAP----- 725
QY
6954 EAPPKAVSKKIPPEEKVPVPEVQKEAPPKVPPEVKVPPEEKVLPKKEAVPPAKGRTVLE 7013
Db

Db 1079 IRVFCNCGHCPSTPATSTATSSPTGGTWTWLTATTTATTTTSTGTAFTSLRTAP 1138
QY 418 AP---TKKPAPT---TPKEPAPTAPKDA---PTTPK---EPAP----- 450
Db 1139 PKVLTTTATPTVTSKATPSSPGTATAPALPASTATPTATGATVPSPSSGLTTWTR 1198
QY 451 ---TPKEPAPTTPKEPSP-----TPKEPAPTTPKAPATTTKBPAP 489
Db 1199 LSQTTTATMTATPSSPTPEAHSTVLTAATTTGATGSAVATPSSPTGTAHTTKVPT 1258
QY 490 TTT---KSAPTTPKEP-----SPTT----- 506
Db 1259 TTTGATPSSPGTALTTPVWISTTTTPTTRGSTVTPSSIPGTTTATVLTATTTTAT 1318
QY 507 ---TKKPAPTTPKEPAP---TPKEPAPTTPKEPAP---TPPKPAPT--- 529
Db 1319 GSWATPSSSTQSGTPSSLTATTTATITATGNTNPSPTGTRPDPVLTATTTATPATSS 1378
QY 530 ---TPKEPAPTTPKEPAPT--- 547
Db 1379 TVTPSSALGTTTPPVNTTATTHGRSLPSSPHVTAWTSATSGTLGTHITEPSGT 1438
QY 548 ---KKPAPT---AKKPAPT---AKKPAPTTP---KETAPTTPK--- 573
Db 1439 SHTPAATGTTQHSALSSPHSSRTTESPPSGTTPGHTTATSTRTATATPSKRTS 1498
QY 574 -LTPTPEKLAPTTP--- 588
Db 1499 TLLPSSPTS-APITTVWVGCEPQAWSEWLDYSYMPGPGSGDFDTYSNRAAGAVCE 1557
QY 589 KP----- 590
Db 1558 QPLGLECRAQOPVPLRELQWECSDFLGLVCRNREOVGKFKMCFNYEIRVFCNNGH 1617
QY 591 APTPEELAPTPEE-----PTTTPKEPAP-----T 617
Db 1618 CBSTATSTATPSSPTGTTWLTQTAATTTATGTAIPSSPTGTAAPPKVLTSQAT 1677
QY 618 TP---KAAPNTPK-----EPAPT---PKP--- 638
Db 1678 TPTATSSKATSSSPRTATLPLVLTATKSTATSTFTPIPSSTGLTGTGSONRPPHMA 1737
QY 639 ---APTTKEPAPTTPKEPAPT--- 660
Db 1738 MSTHPSSTPETHSTVLTATKATYTRATSSMSPSPGTTWLTWLTATTAATTAALPH 1797
QY 661 GTAPT---LKEPAPT--- 674
Db 1798 GTPSPSTGTTWLTPESTTATVTPGTSTATASSTRATAGLKVLTSTATTPTVISSRAT 1857
QY 675 PKK-----PAPKELAPT--- 686
Db 1858 PSSPGTATAPALPASTATPTATSVTAIPSSSLGTAWRLSQTTTPTATMTATPSSSTP 1917
QY 687 ---TKKEPTSTT--- 695
Db 1918 ETVHTSTVLTATTTTRTGSAVATPSSPTGTAHTTKVPTTTTGTATPSSSPGTALTPPV 1977
QY 696 ---SKPAPT--- 703
Db 1978 WISTTTTPTTRGSTVTPSSIPGTTTATVLTATTTTATGSMATPSSSTQSGTPSSLIT 2037
QY 704 ---PKGAPTTPKEPAPT--- 719
Db 2038 TATTITATGNTNPSPTGTTPIPVLTATTTATTAATSSVTPSSALGTHTPPVNTTA 2097
QY 720 ---KPEP---APT-----TPKGTAPT---LKEP 739
Db 2098 TTHGRSLPSSPHVTVPATWTSATSGILGTHITEPSTGTSHTPAATGTTQPSPALSSP 2157
QY 740 AP---TPPKPAPKELAPTTPKGTSTSTSDKPAFTPKET-----AP-----T 779

Db 2158 HPSRTTESPPSP---GTTTPGHTRGTSRTTATATPSTKTRTSTLLPSSPTSAPITTVT 2213
QY 780 TPKEP----- 784
Db 2214 TGCEPOCAWSEWLDYSYMPGPGSGDFDTYSNRAAGAVCEQPLGLECRAQAPGVPLR 2273
QY 785 ---APTPKXPAPTTPPTPP 802
Db 2274 ELQVVECSLDPLVCRNREOVGKFKMCFNYEIRVFCNNGHCPSPATSSATPSSTG 2333
QY 803 TT----- 804
Db 2334 TWTILKLTATTTATTTSTGSTATSSSTQGPAGTPHVSTTATPTVTSSKATPSSPGTA 2393
QY 805 ---SEVSTPTT--- 822
Db 2394 TALPALRSTATTPTATSTAIPTSSSLGTTWRLSQTTTPMATWSTATPSSPTETVHTSTV 2453
QY 823 --- 822
Db 2454 LTTTATTTGATGSAVATPSSPTPGTAHTTKVPTTTTGTGTVTPSSSPGTAARTPPVWISTTTT 2513
QY 823 ---PDEST----- 827
Db 2514 PTTSGSTVTPSSIPGTHPTVLTTPQPVATGSMATPSSSTQSGTPSLITTATTITA 2573
QY 828 ---PELSAEPKALEN----- 841
Db 2574 TGSSTPSSPTGTTPIPELTATTTATPAATSSVTPSSALGTHTPPVNTTATTHGRSL 2633
QY 842 SPKEP-----GVPTTKTPAATKPEMT--- 862
Db 2634 SPSSPHVTAWTSATSGTLGTHITEPSTGTSHTPAATTTGTTTSTPSSALSPHSPSSRT 2693
QY 863 ---TTAKDKTTE----- 871
Db 2694 ESPSPSGTTTPGHTTATSTRTTATATPSTKTRTSTLLPQSTSAPIITVTVTTCPEOCAMSE 2753
QY 872 --- 871
Db 2754 WLDYSYMPGPGSGDFDTYSNRAAGAVCEQPLGLECRAQAPGVPLGELQVVECSLD 2813
QY 872 ---RDLETT 877
Db 2814 FGLVCRNREOVGKFKMCFNYEIRVFCNNGHCPSPATSSSTAMPSSPTGTTWLTWLT 2873
QY 878 PETT---TAAPKMTKETA---TTTEKTTESKI---TATTQV-TSTTT 915
Db 2874 ATTTASTGSTATPSSPTGTAAPPKVLTSPTATTTATSSKATSSSPRTATLPLVLTAT 2933
QY 915 QDT---TPFKITTLKTTTLAPKVTTK-KLITTEIMNKPEE-----TAKPKDRATN 963
Db 2934 KSTATSVTPIPSSILGTTGLPEQTTTPVATMTSTIHPSSPTETHTSTVLTATKATRATS 2993
QY 964 SKAT---TPKPK----- 973
Db 2994 STSTPSSPTGTTWLTWLTATTTAGTPTATSSPTGTTWLTWLTATTTATTTASTGST 3053
QY 974 ---PTKAPKPTS 983
Db 3054 ATLSSTPGTTWLTPESTTATVAPPGSTATASSTQATAGTPHVSTTATTTPTVTSKATP 3113
QY 984 TKPKPT---MPEVRKPKTTPPRKMTSTMPBLNPTS-----RIAEAMLOTTRPNQTP 1033
Db 3114 SSSPGTATAPALRSTATTPTATSTAI-----PSSSLGTTWRLSQTTTPTATMTATP 3168
QY 1034 NSKLVEVNPKSEDAG-----GAETPHMLLRPHVFNPEVTPDMOYLPKVP----- 1079
Db 3169 SS-----TPETVHTSTVLTWLTATTTGATGS-----VATPSSPTGTAHTTKVPTTT 3214
QY 1080 ---NOGILINP 1087
Db 3215 TGFTATPSSSPGTALTP 3231

RESULT 7
T30826
nascent polypeptide-associated complex alpha chain, muscle splice form gp220 - mouse
N:Alternate names: alpha-NAC protein
C:Species: Mus musculus (house mouse)
C:Date: 22-Oct-1999 #sequence_revision 22-Oct-1999 #text_change 09-Jul-2004
C:Accession: T30826
R:Yotov, W.V.; St-Arnaud, R.
Genes Dev. 10, 1763-1772, 1996
A:Title: Differential splicing-in of a proline-rich exon converts alphaNAC into a muscle
A:Reference number: Z20899; MUID:96312450; PMID:8698236
A:Accession: T30826
A:Status: Preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-2187 <YOT>
A:Cross-references: UNIPROT:P70670; EMBL:U48363; NID:g1666688; PID:g1666689; PIDN:AA187
C:Genetics:
A:Gene: Naca
A:Map position: 10
A:Introns: 24/1; 1996/1; 2024/3; 2050/3; 2099/3; 2142/3; 2183/3
A>Note: differential splicing converts alphaNAC into a tissue-specific DNA-binding activ
C:Keywords: alternative splicing; DNA binding; transcription factor

Query Match 16.7%; Score 974.5; DB 2; Length 2187;
Best Local Similarity 21.6%; Pred. No. 4.4e-12;
Matches 364; Conservative 134; Mismatches 419; Indels 765; Gaps 72;

QY 4 KTLPIYLL-----LLSVFIQVSSQDLSSCAGRCBGYSR 40
Db 479 KNLPIALVNGVAPSPAQAAGLPTKRKDTTLOPLAPIALKESPSSQSASL-----EVLSE 533
QY 41 D-----ATCNCYDNCQHYMECCPDFKRVCTAELSCKGRCFSEF 79
Db 534 DVTTKTGGPAPVVRPAIGVAT-----TTSL----- 561
QY 80 RGRCDCAQCKYDKC-----C 97
Db 562 ---RADSPAVIRADSCVSPNTVSQLKRSVTDPAAPRTAKNTAPSTTSPLVPLASEGC 618
QY 98 PYEFCAEHSVSNQSSSSSSSSSTTWIKSSKNSAANFELOKLVKONKQR 157
Db 619 P-----VASSMALSPONASVETAL-----ALSEIERSV----- 648
QY 158 TKKPTKPPV-----VDEAGSLDNGDFKVTTPDTSTTQH----- 193
Db 649 ----PFPDPLAEISFSNARKVDVSHMESSGSSRGQHPDASVTAKTGTVCLADSLDTS 704
QY 194 -----NKVSTSPKAITAKPINRPSLP 215
Db 705 VSASKGSALSGASSPLYFLEVSFLPEAGLAVQGPKGSLNKLSPTPPSKGAPV-PSTGAP 763
QY 216 PMSDTSKETSILVNKETTETKE-----T 239
Db 764 P-----SPKGAPIVPTESSISSQVFAEILPSPQKTPVETASRLISAVOSPKVDPTMSDV 819
QY 240 TTNKQSTSDGKEKTTSAKETOSI-----EKTSAKDLAPTS--KV 277
Db 820 PTPSPKTSATAVPKOTSA--TSLSKSVPAVTSLSPPKAPVAPSNEATVPTIPTSLKNA 877
QY 278 LAKPTPKAETTT-----KG-PALTT-----PKE- 299
Db 878 LAATPKETLATSIPKVTSPSPQKTPKSVSLKGAFAMTSKKATEAASKDVSPSQFPKEV 937
QY 300 -----PTPTPKPASIT-----PKEPTPTTIKSAPTPKE-PAPTTTKSAPTTP 343
Db 938 PLLQHVPTSPSPKSPVSDLSGALTSPPPKGP-PATLAETPTYPKSPKPAASKKTPATP 996
QY 344 K-----EPAPTTKEPAPTTKEPAPTTKEPAP----- 372
Db 997 SPEGVTAVPLEIPPCSKAPKTAAPKESATSSSKRAPKTAATVSKGVTAVPLEISL 1056

QY 373 ---TTTKSAPTTKBPAPTTPKK-PAPTTPKK-----PAPTTPKBPTP-- 411
Db 1057 PLKETSASA--TPGEKSASSPKRSPKTAGPKETPPGCVGTAVPEISLPLPKETPQONATPNE 1114
QY 412 -----TTPKE-----PAPTTKEP-----A 425
Db 1115 SLAASSQKRSPTSVPKETPPGCVGTAMPLEIPSAPOKAPKTAVPKQIPTPEDAVTILLAGS 1174
QY 426 PTTPKBPAPTAAPKBPAPTT-----PAPTTKEP-----KEPAPT 451
Db 1175 PLSPKKASKTAAAPKEAPATPSVGVIAVSGEISPSPKKSTAAAPKENSATLPPKRSPKTA 1234
QY 452 TPKEPAPTTTKE-----PSPPTPK-----PAPTTTKSAPTTTKEPAPTTTKSAPT 498
Db 1235 APKE-TPATSSSEGVTAVPSEISPSPTPASKGVFVTLTPKGAPNALAE-SPASPKKVPKT 1292
QY 499 --PKERS--PTTKEPAPTTKBPAPTTKBPAPTT-----PKEPAPTTKBPAPTT 546
Db 1293 AAPEETSTTPSPQIKPVAKGPKESATPPSKKTKPTAVPKETSAPSEGVAVPLEIPPSPP 1352
QY 547 TKBPAPTAPEK-PAPTTPKETAPTTPKKLTPTTPEKLAPTTPEKAPTTPEELAPTTPEE 605
Db 1353 RKAPKTAAPKETEAPS--PEGATTAPVQI-PPSPKSGSKKAGSKETPTTP-----SPEG 1403
QY 606 PTTTPEEP--APTTPKAAAP-----NTPKEPAPTTKBPAPTTTK 644
Db 1404 VTAAPLEIPISSKKTSMASPKETLVTPSSKKLSQTVGPKETSLEGATAVPLEIPPSHK 1463
QY 645 EPAPTTKBPAPTTKGTAPTTLKEPAPTTKBPAPKAPKELAP-----TTTKEPTSTTDKPA 700
Db 1464 APKTVDPKQVLPSPKADAPTTLAE-SFSPKPK-APKTAAPSERVTVTP-----EKPA 1516
QY 701 PTTPKGTAPTTKBPAPTTKPEPA-----PTTPKGTAPTTLKEPAPTTKPKKAPKE 751
Db 1517 -TPQASGTTASKVPVPAETQEVAVSRETFTVPAVPVKNQPSHKKTSKTIELKEAPAT 1575
QY 752 LAPTTTKGPTSTSDPAPTTPKETAPTTKBPAPTTKBPAPTTKBPAPTTTETPPTTSEVSTPT 811
Db 1576 LPESPTKSPKPSKK-----APTSAP-----KEFPASSEIKPVTSLSAQT-- 1617
QY 812 TTKEPTTIHKSPEDESTE--LSAEPKPALENSPKBPVPTTKTPAATKEPMTTAKDKT 869
Db 1618 --APSLQKAPSTTIPKENLAA--PAVLPSVSSKSPAAP-ARASASLSLSP--ATAAPQT 1667
QY 870 TERDLRTP-----ETTTA-----APKMTKETATTTTEKTESKITATTQVTS 912
Db 1668 APKEATTIPSCKAAATETPIETSTAPLEGAPKETSETSV-----SKVLSSPPPKA 1720
QY 913 TTTQDTTPEFKITL---KTTILAPKVTTTKITTTTEIMNKPEETAAPKDRATNSKAT 968
Db 1721 SSKRGASTLPATLPSLKEASVLSPTATSSGK-----DSHISPVSDACSTGTT 1769
QY 969 PKQKPTKAPKK-----PTSTK 985
Db 1770 --PQASEKLPKXGPTAFTTEMLAAPAPESALAITAPIQKSPGANSNSASSPKCPDPSSKK 1827
QY 986 KPCTMPR--VRKPKTTTPPKMTSTWPELNPTSRIAEAMLOTTTRNQ----- 1031
Db 1828 DTGCLPSAVALAPQTVPEK-----DTSKAIETLLVSPAKGSDCLHSPKGPVGSQ 1877
QY 1032 -----TPNSKLVENPKNSEADAGAEGETP 1055
Db 1878 VATPLAFTSDKVPPEAVSAPKAPAPASLTLAPSPVAPLPKQPLLESAPGVSLESP 1937
QY 1056 HMLLRPHVFMPEVTPDMYDLPV-----PNQGIILN-----P 1087
Db 1938 SKL-----PVAEEDELPPILPPEAVSGGEPFQDILVNMPPAKPAGTAPAPAPSAKQP 1989
QY 1088 ML 1089
Db 1990 VL 1991

| | | | | |
|--|-----|--|---|------|
| RESULT 8 | Db | 221 | STTSTSSSSSTTS | 235 |
| S48478 | Qy | 177 | DNGDFKVTTPDSTTHQNKVSTPKITTAKPINP | 228 |
| Glucan 1,4-alpha-glucosidase (EC 3.2.1.3) - yeast (Saccharomyces cerevisiae) | Db | 236 | -----STSSSTTSSSSSTSSSTTAPATPTTSCTEKPTPTTSCTEKPTPP | 288 |
| N/Alternate names: extracellular glucosylase; mucin-like protein MUC1; protein YI019c | Qy | 229 | NKETTVEYKETTNNKQSTDKKETTSAKETQSTIEKSAKDLAPTSLKVLAKPTKAEPT | 288 |
| C/Species: Saccharomyces cerevisiae | Db | 289 | HHDTTPCTKTKKTTTSK-TCT---KTTTPVPTPS-SSITSSSAPV | 337 |
| C/Date: 10-Sep-1999 #sequence revision | Qy | 289 | TKGPALTTTPE | 336 |
| C/Accession: S48478; A26877; E26877; S27281; J06123 | Db | 338 | ESSSAPVSTSSSTSSSAPVPTSSSTSSSAPVSTSSSAPVSTSSSAPVPT | 397 |
| R/Rowley, K. | Qy | 337 | KGAPTTTKEPAP | 372 |
| submitted to the EMBL Data Library, October 1994 | Db | 398 | PSSSTSSSAPVSTSSSTSSSAPVSTSSSAPVSTSSSAPVSTSSSAPVPT | 457 |
| A/Reference number: S48478 | Qy | 373 | TITKGAPTTKEPAPTTPKKPAPTTPKAPPTTKEPTP | 423 |
| A/Accession: S48478 | Db | 458 | VTPSSSTSSSAPV | 509 |
| A/Molecule type: DNA | Qy | 424 | ---PAP--- | 462 |
| A/Residues: 1-1367 <ROW> | Db | 510 | APVPTPSSSTSSSAP-APTSSSTSSSAPVSTSSSAPVPTSSSTSSSAPVPT | 568 |
| A/Cross-references: UNIPROT:P08640; GB:Z47047; EMBL:Z38061; NID:g603997; PID:g763364; GS | Qy | 463 | BSPTTKEPAPTTPKSAPTTTPKAPTTTSAPTT | 512 |
| R/Yamashita, I.; Nakamura, M.; Fukui, S. | Db | 569 | VTSSTSSSAPVPTSSSTSSSAPVPTSSSTSSSAPVPTSSSTSSSAPVPT | 628 |
| J. Bacteriol. 169, 2142-2149, 1987 | Qy | 513 | ---TPKEPAPTTPKAPTTTPKAPTTTTPKAPTTTPKAPTTTPKAPTT | 570 |
| A/Title: Gene fusion is a possible mechanism underlying the evolution of STA1. | Db | 629 | SITESSSAPVPTSSSTSSSAPVPT | 684 |
| A/Reference number: A91831; MUID:87194600; PMID:3106330 | Qy | 571 | PKLPTTPEKLAPTTPKAPTTPEELAPTTPBPTP | 626 |
| A/Accession: A26877 | Db | 685 | ---SSTSSSAPV--- | 732 |
| A/Molecule type: DNA | Qy | 627 | PKEPAP | 680 |
| A/Residues: 762-1331 <YA2> | Db | 733 | SSAPVPTSSSTSSSAPV | 788 |
| A/Cross-references: EMBL:M16165; NID:g172523; PIDN:AAA35015.1; PID:g172526 | Qy | 681 | KELAPTTPKEPTSSDK | 725 |
| R/Pardo, J.M.; Ianez, E.; Zalacain, M.; Claros, M.G.; Jimenez, A. | Db | 789 | SSSAPVPT--PSSSTSSSAPVPT--TPSSSNTSSAPSTPSSSTSSSAPVPTPSS | 845 |
| FEBS Lett. 239, 179-184, 1988 | Qy | 726 | TPKGTAP | 776 |
| A/Title: Similar short elements in the 5' regions of the STA2 and SGA genes from Sacchar | Db | 846 | TTSSSAPVSSSTSSSAPV | 898 |
| A/Reference number: S27281; MUID:89031230; PMID:3141213 | Qy | 777 | APTTPKEPAPTTPKAPTTTPE | 834 |
| A/Accession: S27281 | Db | 899 | TPSSSKYFGSQTETSVSSTTETTVPTKTTTSTVPTTITTTVCSGTNSAGETTS | 958 |
| A/Molecule type: DNA | Qy | 835 | TPKALENSPKGVPVTKTTPATKPEMTT | 886 |
| A/Residues: 1-31 <PAR> | Db | 959 | SPKVTIT | 1013 |
| A/Cross-references: EMBL:X13857; NID:g4551; PIDN:CAA32069.1; PID:g4552 | Qy | 887 | MTKEATTTEKTESKITATTQVST | 931 |
| R/Lambrechts, M.G.; Bauer, F.F.; Marmur, J.; Pretorius, I.S. | Db | 1014 | STSPSETASESTTTPPTVTTVSTTWTTEYSTKPGGEITTTFTVKIPTTYLTTI | 1073 |
| Proc. Natl. Acad. Sci. U.S.A. 93, 8419-8424, 1996 | Qy | 932 | A--PKVTT--TKKTIITTEI--MKNPET--AKPK | 958 |
| A/Title: Muc1, a mucin-like protein that is regulated by Mss10, is critical for pseudohy | Db | 1074 | APTPSVTTVPTPTTITTTVCSTGTSAGETTSKPKVTITTTVPGSTGTEVTEATT | 1133 |
| A/Reference number: J06123; MUID:96323237; PMID:8710886 | Qy | 959 | -----DRATNS--KATT-----PKQKPTKAPKP | 981 |
| A/Accession: J06123 | Db | 1134 | LVTAVTTTAVTTTSSSTGTSAGTTTGYTKSVPTTWTTLAPVPTPATNAVPTIT | 1193 |
| A/Status: nucleic acid sequence not shown | Qy | 982 | -----TSYKKPKTM-----PRVRKPTTTPKMTSTMEIN | 1013 |
| A/Molecule type: DNA | Db | | | |
| A/Residues: 1-1367 <LAM> | | | | |
| A/Cross-references: GB:U30626; NID:g1304386; PIDN:AAC49609.1; PID:g1304387 | | | | |
| C/Genetics: | | | | |
| A/Gene: SGD:MUC1; STA2; MAL5; DEX2; SGD:S0001458 | | | | |
| A/Cross-references: MIPS:YI019c; SGD:S0001458 | | | | |
| A/Map position: 9R | | | | |
| C/Superfamily: yeast glucan 1,4-alpha-glucosidase homolog; glucan 1,4-alpha-glucosidase | | | | |
| C/Keywords: glycosidase; hydrolase; polysaccharide degradation; transmembrane protein | | | | |
| F/5-21/Domain: transmembrane #status predicted <TM1> | | | | |
| F/1350-1366/Domain: transmembrane #status predicted <TM2> | | | | |
| Query Match | | | | |
| Best Local Similarity | | | | |
| Matches 347; Conservative 110; Mismatches 456; Indels 547; Gaps 65; | | | | |
| Qy | 13 | LLSVFVLQVSSQDLSACGRCGSGYGRDTC | -----NCDVN----- | 49 |
| Db | 6 | LLAVLVLSLNSALGPTALVPRGSEGTSCNSIVNGCNLDFNMHMDQNTMQVTLQV | 65 | |
| Qy | 50 | ----- | ----- | 49 |
| Db | 66 | TSVSWQDNTYQITIHVKGENIDLVLSLKIIGVTGPKGTQVLGYENNTYLDNPTD | 125 | |
| Qy | 50 | -----CQHYNECCDFKRVCTAELSCRCFE----- | 76 | |
| Db | 126 | FTATFEVATQDQVNSQVWM---PNFQI | -----FEYLSQAQYASQWNGTT | 171 |
| Qy | 77 | SFERGEC | -----DCDAOCKKYDKCCPDYSGFCABEHSVSNQES | 116 |
| Db | 172 | SFDLSTGCNNYDNGHQSQTDFPGFYWNIDCDNNC | -----GGTKSSSTTSSES | 220 |
| Qy | 117 | SSSSSSSSSTIWKIKSSNSAANRELQKLKVKUNKQRTKKKTPKPPVDEAGSGL | 176 | |

Db 1645 EGSFKCEHSGFGEDGKIKCTNPLERSCEBDEKFCGRVDHVHSLSVIYNGSLSSVCECE 1704
QY 432 -----PA----- 433
Db 1705 PGRPEKESNCVDIDECEBNNCPASAVCNTGSEYCECAGYEGGVCVTDIDEC 1764
QY 434 -----PTAPK----- 438
Db 1765 DRGAGCDNMAGINRMGCGCKMAGYTGCGATCIKIBEEPCKDKTACTDEWSLCELE 1824
QY 439 KPAPTTPKPAP----- 450
Db 1825 KKQCTVDEEVPCGACLPQHHPINGTCOSLOISGLCAQKNDCKHAECIDIHPSHFCS 1884
QY 451 -----TPKE 455
Db 1885 CPDGFIDGMI CDDVDECNAGCDDNTKCENTTIGSFNCVLEGFKKVDKCVVDEKKQ 1944
QY 456 P-----APTTKEPSPTTPKEPAPTTTKSAPTTPKEPAPTTTKSAPTTPKEPAP 503
Db 1945 PNRKEIIEBENSNSNGQKPTTKGIVSSTSATSSSESTTAEPHVTTISSTSTKDMT 2004
QY 504 PTTKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTP 563
Db 2005 SSKSPENVMTSSSEPSVSSSKSTTASSTTVSSSPSS-----SSAPLTSSPATT 2057
QY 564 KETAPTPKLTPTTPE-----KLAPTPE-----KPAPTTPELAPTTPEPTP 608
Db 2058 TEV-ITESSVKSTTPKESSEITVKLSSKSPETVSSKSPSTPS-----TTSQSVTS 2111
QY 609 TTPPEPAPTTPKAAAPNTPKP-----APTTKEPAPTTTPKEPAPTTPKET-----AP 656
Db 2112 TVPETSVKSLSEAPVTSSTEVHTSETKPSLSASSTGDTNSTTSPSSLASVKST 2171
QY 657 TTPKGT-----APTTLKEPAPTTPKAPKELAPT-----TTKEP----- 691
Db 2172 SAPEGTSAPVAPKLSLSPDV-SQPSTKTFDATESSTVQASESSSGTSVKSTSEPSHV 2230
QY 692 --TSTSDKAPTTTPKGAPTTTPKEPAPTTTPKEPAPTTPKG----- 730
Db 2231 TKLSITSSNPSSVPVTSKSTPTTPEST--EQPTSTTPSQSLTPMNSNSELVTTSEPH 2288
QY 731 -----TAPTLKEPAPT-TPKFPAPKEL--APTTPKGT--APTTP-----STT- 764
Db 2289 VLSSLSLSPDVQSSTTPNNLSSESTVETPKTSSEVSLNSEPSTTEAPTTLSPDILSTT 2348
QY 765 -----SDKP-----APTTP----- 773
Db 2349 NNLQSSTVSTEDRSEISENSEKPTSAPELVTSVTHVASSPDVPTSESPDLTGSS 2408
QY 774 -----KETAPTP-----KEPAPTTPKP-----APTTP----- 798
Db 2409 TENIPEASKQIISSTPTDITTAEEPTKSTNSPDLSSTNSVLSESTTPESSSKSPV 2468
QY 799 -----TPPTTSEVSTPTTKEPTT----- 818
Db 2469 SSSTEGISVVTSTEFKVPKESTISSVLLEDLTTPSPILEETTTASESTSEPLTSLTV 2528
QY 819 ---IHK--SPDESTPELSAEPKALENSP-KEPG-----VPTT----- 851
Db 2529 SVRIHELITSENVPKESSTTSSESKPQEPAGILTSVTVPTSSVLITASEIEAI 2588
QY 852 -----KTPAATKPM-----TTAKDKTTERDLTTTETTTAAKPMKETAFT 894
Db 2589 TSNTPFKQGRTPITTPSKLVKSTTSPSTVTSSEPSSESTKRTTVSTTVSTTPTTEETTS 2648
QY 895 -----TEKTSKLTATTQVTS-----TTQDTTFPKI 923
Db 2649 ESLIITAAPSKPTSESTTSSSEAPTTPAKTSETKPSNVSTGRKSTENVSTISOSGLES 2708
QY 924 TLLKTT-----TLAPKVTTTKTIPTT----- 945

Db 2709 STMSTSSSEPTNAPAVTVSSSEASSTLEENSTSTSEASVKSLSLSPFSEITSAVT 2768
QY 946 -----EINNKPEE----- 953
Db 2769 VSSRAPAEITWSSSHREISTVSSSESEPEIPLSTTVSPNVVTASSIPSEEPILSSVTSS 2828
QY 954 -----TAKPKD-----RATNSKA-----TTPKPKOKPT 975
Db 2829 STPRVLITCTPDDLIVSVTVPSHGRRQNTITASSVPSNSTSPIILPSESLLTTPQPPPT 2888
QY 976 KAPKPKSTTKPKTMRVRKPK--TTPTP-----RKMTSTMPELNPTS 1016
Db 2889 TTTAKPATTSGKGPFSIQPPAEMFTTPAPPPNSNGYGEETNQEEQVIST-----TT 2942
QY 1017 RIAEAMLOT-----TTRNQTPNSKLVEVNPKSSEDAGGA 1050
Db 2943 TEAPSLCSTVTCHSLATCEOSTGVCICRDGFIGDGTGTTACSKKSTADICISLPSLCADKAC 3002
QY 1051 EGET-----PH--MLLRPHVFMPE 1067
Db 3003 DNSTRCECDAGVIGDGYVCSPPQDCVLRDNLCSPE 3039
RESULT 10
S49915
extensin-like protein - maize
C:Species: Zea mays (maize)
C:Date: 05-Mar-1995 #sequence_revision 12-May-1995 #text_change 09-Jul-2004
C:Accession: S49915
R:Rubinstein, A.L.; Broadwater, A.H.; Lowrey, K.; Bedinger, P.A.
submitted to the EMBL Data Library, June 1994
A:Description: Pex genes: pollen-specific genes with extensin-like domains.
A:Reference number: S49915
A:Accession: S49915
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-1188 <RUB>
A:Cross-references: UNIPROT:Q41805; EMBL:Z34465; NID:g600117; PID:CAA84230.1; PID:g60011
Query Match 15.3%; Score 888.9; DB 2; Length 1188;
Best Local Similarity 20.1%; Pred. No. 9.2e-11;
Matches 289; Conservative 87; Mismatches 423; Indels 641; Gaps 54;
QY 10 LLLL--LSVFTVQVSSQDLSSCA-----GRCGE----- 36
Db 17 LLLLAACLSACSQAVTSAEASVIAHQQLLAMKEAGGEGAGDLPADEFDDRVGAANFN 76
QY 37 -----GYSRDATCN-----CDYNCHYMECCPDFKRVCTAEL 68
Db 77 PLRRAYIALQAWHRAFYSDPKGY-----TANWGEDVCKYNGVICTEALDDPKITVAGI 132
QY 69 SCKGR----- 73
Db 133 DLNGADIAGVLPPELGLLTDLAFFHINTNRFGLIPKMSRLSLLHEPDSNNRFVGVFP 192
QY 74 --CFES-----FERGECDDA----- 88
Db 193 YVLEMSVLKYLDRFNDPEGELPALFDK----DLDAIFVNTNRFVGPENLGNSTAS 248
QY 89 -----QC-----KKYDKCCP-----DYESFCAEE 107
Db 249 VIVFANNAFVGPICKSIGRMVKTLDIEIPLNKLDCCLPLEMGLLVNTVID----- 300
QY 108 HSVENQESSSSSSSSSSSTTWIKSSKN-----S 138
Db 301 --VSGNMLVGTIPEQLSNIAKLEQLDVSERNVFTGIVHESICELPALVNFSAFNFFNSEA 358
QY 139 AANRELQKLVKDKNKN-----RTKKK-----PTKPEPVD 170
Db 359 AVCMPSKALVNLDRDNCGLRALRPAQKTALOCAPVLRPVDSCXKVCAGVPTP----- 412
QY 171 EAGSGLDNGDFKVTTPDTSTTOHNVKSTSPKITTAKPINRPSLPSPSDTSKETS LVNK 230

Db 610 STTB--BTITDEPTTT---ESSTTGKATPPLSTTSEBTITELK---ITTE----- 655
 QY 682 ELAPTTTKEPTSTSDKAP---TTPKGATPTTKEPAPPTTKEPAPPTTKEPAPPTTLK 737
 Db 656 --GSTTTEPTTALFAEASGIIITDEBTSTSTTPEITSTKE--IVTESA:QTQSVS 711
 QY 738 EPAPTP---KXPAKELAPTTTKGP----- 760
 Db 712 VVESTPQLPBRKAIUNVKHNLVLEKXRLKKEKESSTTGSDESSTTVVAENID 771
 QY 761 --TSTSDKAPPTTKEPAPPTTKEPAPPTTKEPAPPTTKEPAPPTTKEPAPPTTKE 815
 Db 772 EVTTTEKEKVQTTPTTTEKSTQSETTTTTTEKTTSTTTEKTTSTTTEKTTSTTTEKTT 831
 QY 816 PTTHKSDESTPELSAETPKALENSPKPGVPTTKTPAATKPEMTTAKDKTTERDLR 875
 Db 832 PST---TEST-----TVDTSATTBESSTAETTTISAE-- 862
 QY 876 TTPETTTA-----APKMTKETATTE 896
 Db 863 -TSETTSESAAFITGESPENTALOSSQKSEENESSAEKPGARRDFVPKHKTTVKPAE 921
 QY 897 KTTESKITATTO-VTST---TQDTTTPFKITLTKTTTLAPKV---TTKKTITTTIMN 949
 Db 922 TTSVAASATTTTETPTTTEKSTTLETTTPEATLNEVTGPAFTGAPVDETINTLELLS 981
 QY 950 KPEETAKPKDRATNSKATTPKPO-----KPTKAPKPTSTKPKTMPR 992
 Db 982 K-----INNTQISQPKPTDISKTDLSSLLISGLIGSFTKAPMAPTI----- 1022
 QY 993 VAKPKPTTTPRKMTSTMPLN-----PTSRIAEAMLOTTRPNOTPNKSLVEV 1040
 Db 1023 ---HTTDAAPVATATEASLNDGSKTIIDEAQPTDEIRRA-----OPTN----- 1063
 QY 1041 NPKSEDAGGABGETHMLLRPHVFMPEVTPMDVLPVYPNGI 1083
 Db 1064 -----EMDKEMEFKRIREQRI 1080

 RESULT 12
 T29757
 Protein UNC-89 - Caenorhabditis elegans
 C:Species: Caenorhabditis elegans
 C>Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 03-Dec-1999
 C:Accession: T29757
 R:Du, Z.; Le, T.T.; Wilson, R.
 Submitted to the EMBL Data Library, May 1997
 A:Description: The sequence of C. elegans cosmid C09D1.
 A:Reference number: Z20679
 A:Accession: T29757
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 1-6642 <DUZ>
 A:Cross-references: EMBL:AF003131; PIDN:AAB54132.1; GSPDB:GN00019; CESP:unc-89
 A:Experimental source: strain Bristol N2; clone C09D1
 C:Genetics:
 A:Gene: CESP:unc-89
 A:Map position: 1
 A:Introns: 17/2; 108/3; 154/2; 211/2; 265/3; 326/2; 352/3; 426/2; 454/1; 500/1; 537/1; 6
 /3; 5917/1; 6027/1; 6061/3; 6153/2; 6515/1; 6552/3; 6609/1

 Query Match 14.8%; Score 862.5; DB 2; Length 6642;
 Best Local Similarity 13.9%; Pred No. 3.3e-09;
 Matches 397; Conservative 146; Mismatches 385; Indels 1925; Gaps 96;

 QY 34 CCEGYSRDATCNC-----DYNCOH----- 52
 Db 659 CKEGDVDFCEVEGWPPPELVWLDQPLRPSHDFRLQYDQGTAKLEIRDAQPDGTGVY 718
 QY 53 -----YMECCPO-----FKRYCTAE-- 67
 Db 719 TVKIQNEGSGISBKAELFVQADPDKNHVAPEFQATIEVEEC--DEGEVRFKSVITGPN 776

QY 68 -----LSCKGRCFESPERGRCDC 86
 Db 777 PELIWFINKPLSEBKVFISEGDICILITIKDVTTRHFDGMVTCQS-----NRLGSASC 831
 QY 87 DAQCK-----KYDKCCPD-----YE-----SFAEE----- 107
 Db 832 DGRKLVVPAPPTFNKPLEDKTQEKSTVVPFVDSGVWPEPTLTFLTCLGKELKNGEGV 891
 QY 108 -----HSYSEN-----QESSSSSS----- 122
 Db 892 EIVGHDGFYRISIPNTSMKDHGEIVAKAQNDEHGAESRARLTVQEESRSAPFLKD 951
 QY 123 -----SSSSSTIWKIKSK----- 136
 Db 952 IEDQTVKTGEFAVFETTVRGNPNPEVTFINGHKMDQSGPGVKIEAHNHDHKLITDSAQY 1011
 QY 137 -----NSAANRELQKLKVKONKKKRTTKKP----- 162
 Db 1012 AGTVLCRAENAVGRFETKARLVVLAPEK--OKKPKFVFIILVDKTETVDNTVVEFVVE 1068
 QY 163 -TPKPPV-----VDEAG-----SGLDNGD 180
 Db 1069 GEPKPTVWYLKGEELKQSDRVEIREFDGSIKISIKNIKIEDAGELRAVATNSESDET 1128
 QY 181 FKVT----- 184
 Db 1129 AKLTVOKPPFAPEFDLPRVSLTVEKGEAVFSAHAFGIPLPTYEWSVNGRKYRQDQEGAR 1188
 QY 185 -----TPDTST-----TQHNKVS----- 197
 Db 1189 VTRDESTVDSAILITIDTAYYSEVNNHLTISVVAENTLGAETGAQLIEPKESVWVEK 1248
 QY 198 -----TSPKITTAKPINRPSLPNSDTSKETSLSLVNKTETVETKET 239
 Db 1249 QDLSSSEVQKIAQVKEASPEATIT-----ITMETSLSITKTITMTSTTEV 1294
 QY 240 TIT----- 242
 Db 1295 TSTVGGVTVETKESESESATTVIGGGGGVTEGSIKIEVSVKTDQTDVREGTPRRR 1354
 QY 243 -----NKOTSDGKEKTTSAKETOSIEKTSAKDLAPTSKVLAKPTP 283
 Db 1355 VSPABEELPKEVIDSDRKKKSPSPDKKEK-----SPEKTEKSPASPTKKT----- 1400
 QY 284 KAEPTTKGALITTPKEPTTTP---KEPASTTPKEPT-----PTIKSAPTTP--- 328
 Db 1401 -----GEEVKSPEKSPASPTKKEKSPAAEVEKSPTKKEKSPSPSPSPSPSPSPSP 1453
 QY 329 -----KEPAPTTKSPAPTTPKEPAPTTP---KEPAPTTP---KEPAPTTP--- 367
 Db 1454 TGDEVKEKSP-----PKSP---TKKEKSPKEDVKSPVKKEKSPDANIVEVSSE 1501
 QY 368 -----KEPAPTTKSPAPTTP---KEPAPTTPKSPAPTTP--- 382
 Db 1502 TTTEKTETTTMTTMTTTHESSESRTSVKKEKTPKDEKPKSPTKKDKSPKESKITESIKSPV 1561
 QY 383 -KEPAP-TTPKPKPAPTTPKEPAPTTPKEPTTPKPE-PAPTTPKEPAP-----TTPKE 431
 Db 1562 KKEKSPKVEKPEKSPKKEKPEKSPKSPKSENEKSPKSPKSENEKSPKSPKSENEKSPK 1621
 QY 432 PAP-TAPKPKPAPTTPKEPAPTTPKSPAPTTPKPEPAPTTPKPEPAPTTPKSPAPTTP 490
 Db 1622 KSPEKADDPKSPKKE-----KSPKESATSDVKSPKKEKSPKKEKSPKKEKSPK 1670
 QY 491 TTKSAPT-----TPKEPSPTTKE-PAPTTPKEPAP-----TTPKPKPAPTTP 531
 Db 1671 KKESSPTKTTDDDEVKSPKKEKSPKQVEKPKSPKKEKSPKSENEKSPKSENEKSPK 1730
 QY 532 KE-PAPTTPKEPAP-----TTTKK---PAPTAPKEPAPTTPKETAAPT----- 569
 Db 1731 EEKPKSPKKEKSPKSENEKSPKSENEKSPKSENEKSPKSENEKSPKSENEKSPK 1790
 QY 570 -----TPKKL-----TPTTPEKAPPTTPKEPAPTTPPEEPTPTTP----- 611

Db 1791 PTKKSPKPKVEKASPTKKK-----TPEKSA---AEBLKSPTKKESPPSKTKTGDE 1843
QY 612 -----EPAPPTPKAAAPNPKBPAPPTPKBPAPPTPKBPAP----- 648
Db 1844 SKEKSPKPEPKSPPTPKSPKPGSPK-----KKSKSPEAKPPAPKLTRDLKLQTVNK 1898
QY 649 ----- 648
Db 1899 TDLAHFVVVVEHATECKWFLDGEKITTAGVTVSKDDQPEPRCSIDTTFMFGSGTVSVVAS 1958
QY 649 -----TTPKET----- 654
Db 1959 NAAGSVETKELKVLKLETPKTKPKPTDKLRDMVETKGDVQMDVIALHSPLYKYQNGN 2018
QY 655 -----AFTPKGTAPITL 667
Db 2019 LLEDGKNGVTIKNENKSSLIIPNAQDSGKIIVASNEVGSSSSAQLTVNPPSTPIVV 2078
QY 668 KEPAPPTPKKAPKAPKELAPTTTKEPTSTTSDDKAPPT----- 703
Db 2079 DGPKSVTIKE-----TETAEFKATISGFPAPTIVKWTINEKIVEESRTIITIKTEDVY 2130
QY 704 -----PKGTAPTTPK-----BPAPTPK----- 721
Db 2131 TLKISNAKIEQGTGVKVTQAQNSAGODSKOALDKVBNVKAPKPKGSQLTKVXADEGEPLRW 2190
QY 722 -----EPAPPTP-----KGTAPTTTKEPAP----- 741
Db 2191 NLELDGSPGCTEWSLLNGOPLTKSDTVQVVDHGDGTVHTIAEAKFEMSGTLTAKAKNA 2250
QY 742 -----TPPK-----KPAPK----- 750
Db 2251 AGCETSATKYTVNGGNKKEPEFVQAPQNHETLEESVKFSAIVTGKPMNVWYLNKKLI 2310
QY 751 -----ELAPTT 756
Db 2311 QSEEVKVVVHETKTSIRIQKPLMEHNGTIRVEAENVSGKVQATAQLKVDKKTVPKET 2370
QY 757 T-----KGPTSTTSDDK-----APPT 772
Db 2371 TNMDRQKGEDEVKFTANVGVPEPSVAWTLNGEPVSKHPNITVTDKGEHTIETSAVT 2430
QY 773 PKET-----APTTPK----- 782
Db 2431 PEQAGELSCBATNPGSKKRDVQLAVKVGADPTFAKNLEDRLITEGELTMDAKLNIUVK 2490
QY 783 ----- 782
Db 2491 PKKTIWLKGVETSDGHVKIVEEDGSLKLSILQTKLEDKGRITIKAESEFVGAESA 2550
QY 783 -----BPAPT----- 788
Db 2551 SLGVVKGPRWAKPAFQSDIAPINLTGDTLECKLLITGPTPFVKWYIGTQLVCATEDTE 2610
QY 789 -----PKK-----PAPPTPET----- 799
Db 2611 ISNANGVYTMKIHGVTADMTGKIKVAYNKAGEVSTEGPLKVAPIPVEFEISLCDATCR 2670
QY 800 -----PPPT-----SEVSTPTTKETPT----- 818
Db 2671 EGDTLKRAVLGPEPEPVWSWYNGKLEBSONIKIHSEKGTVTVIKDIICDYSQQVVC 2730
QY 819 -----IHK----- 821
Db 2731 EAINEXKATSEATLLVLRGEPDFLEWLSNVRARTGTQVVKVVFVTDGPKESLWYIN 2790
QY 822 -----SPD----- 824
Db 2791 NKEILNSDLTYITDDKTSLTINSFNPDVHVGELICKAENDAGEVCTANMITYTDMF 2850
QY 825 ESTPELSAE-----PTPKALENSPK-----E 845

Db 2851 SESESEAAQAEFFVGDLDLDESLREEMHRTPTVM---APKFTIKTKDKAKKXGHSVFE 2907
QY 846 PGVPTTK----- 852
Db 2908 CVVPDTHGVCCKWLKDGKEIELIARIVQTRTQPEGHIHQELVLDNVTPEDAGKYTCIVE 2967
QY 853 -----TPAAT-----KPEMTTTAKDKTTTBERDLRT-----PPTTT 882
Db 2968 NTAGKDCETATLTVIESLEKKSEKKAPEFIVALQDKTKTKSEKVLCKVIGEPKPKVSW 3027
QY 883 AAPKWTETATTTTETKT-TESKITATTTO---VTSITTTODTTPPKITTL---KTTTLAP 933
Db 3028 LHDNVSEKPNSEKTIITQESITVESVEGVERVTISSE-----LSHQGYTCIAE 3078
QY 934 KVTITTKT-----ITTEIMNKP-----BETAK 956
Db 3079 NTEGTSKTEAFLTVQGEAPVFTKELQNKELSIGELKVLSCSVKSGSPOPHVDYFSFSETTK 3138
QY 957 PKDR-----ATNSKATTPKPKPT---KA 977
Db 3139 VETKITSSSRIAEHDQTNTHRMVISOITKEDIVSKAIATNSIGTATSTSKITTKVEA 3198
QY 978 P-----KKPTSTK-----KP-----KTMPR----- 992
Db 3199 PVFEQGLKKTIVKEKEIKMEVKVGSAPDVENFKDDKPVSEDNHEMKKNPETGVFTLV 3258
QY 993 VRKPKTT-----PT-PRKMTSTWPELNPTSRIAEA 1021
Db 3259 VKQAATTAGKYTAKASNPAAGTAESSAAEVQTSLEKPTFVRELVTTEVKINET-----A 3313
QY 1022 MLQTTTR-----PNOTPNSKLV-----EV 1040
Db 3314 TLSVTVKGVDPSPVSEWMLKDGQPVQTDSSHVIKAVEGSGSYSIITIKDARLEDSGKVACRAT 3373
QY 1041 NPKSEDAAGAGETPHMLLR---PHVFMPEVTP 1070
Db 3374 NP-----AGEAKTEANFAVVKLVPPPEFVKLSP 3402
RESULT 13
B48666
cell proliferation antigen Ki-67, short form - human
C:Species: Homo sapiens (man)
C:Date: 03-May-1994 #sequence_revision 03-May-1994 #text_change 15-Mar-2004
C:Accession: B48666
R:Schluster, C.; Duchrow, M.; Wohlenberg, C.; Becker, M.H.G.; Key, G.; Flad, H.D.; Gerde
J. Cell Biol. 123, 513-522, 1993
A:Title: The cell proliferation-associated antigen of antibody Ki-67: a very large, ubiq
ins.
A:Reference number: A48666; MUID:94043435; PMID:8227122
A:Accession: B48666
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-2897 <SCH>
A:Cross-references: ENBL:X65551
C:Keywords: alternative splicing; cell cycle control; nucleus; tandem repeat
F:29-91/Domain: kinase interaction domain homology <KIH>
Query Match 14.4%; Score 839.5; DB 2; Length 2897;
Best Local Similarity 15.5%; Pred No. 3e-09;
Matches 408; Conservative 140; Mismatches 445; Indels 1635; Gaps 99;
QY 4 KTLPTYLLLLLSVFIQVSSQDLSSCAGRCGEGYSDATC----- 44
Db 241 KTA-----ASSKSGQTEVPKRGGE---RVATCLQKRVISRSQSDILOMIC 284
QY 45 -----N 45
Db 285 SKRRSGSEANLIVAKSWADVVKLGAKOTQVKIKHGQORSNKNKQRRPATPKPKVGEVH 344
QY 46 CDYNGQHYMECCP-----DFKR----- 62
Db 345 SQFSTGHANSPTIIGKAHTEKHVHPARVYVRLNFIQNMDFKEDLSGIAEMFKTPV 404


```

QY 555 PKRAP--TTPKET--APTTP-----KKLTPTT----- 578
Db 2103 CKSPPPESMDTPTSTRRPFKTPGLKRDIVIELSALKQLTQTHDKVFGDEKGINVERE 2162
QY 579 --PEKLAP--TTPKPAPTTP-----EELA-----PTTPEETPTTPE 612
Db 2163 TAKQKLDPAASVTGSKRQPTPKGAQPLEDLAGLKELFPQVCTDPTTHEKTKTACR 2222
QY 613 EPAP-----TTPKAAA----- 623
Db 2223 SPQDPVGTPTIFKPSKRSRLKADVEESLALRKRTPSVGKAMDTKPPAGGDEKMKAF 2282
QY 624 -----PNTKPEPA-----PTTKPEPATTP 643
Db 2283 MGTVPQKLDLGNLPGSKRWPTPKRAQALDLAGFKELFQTPGTDKPTTDEXTKIAC 2342
QY 644 KEPAP--TTPKET-----APTTPK----- 660
Db 2343 KSPQDPVDTASTKQRPKNLRKADVEEFLALRKRTPSAGKAMDTPKAVSDEKNIT 2402
QY 661 -----GTAPTLKEPATTPPKPAP-----KELAPTTTKPTSTSDX-- 698
Db 2403 FVETPVQKLDLGNLPGSKRC--PQTPKRAEALDLVGFELFQTPGCHTEESMTDKIT 2460
QY 699 -----PAP-----TTPKGTAFT-- 710
Db 2461 EVSCKSPQESFKTSRSKQBLKPLVKVDKMBPLAVSKLRTSGTQTQHTPTGDSK 2520
QY 711 -----TPK--BPAPT-----TPKEPA----- 724
Db 2521 SIAFKESPKQILDPAASVTGSRRLRTRKEKARALEDLVDFKELFSPAGHTEESMTIDK 2580
QY 725 -----PTTPKGTAFTLKEPATTPPKPAPKELAP-----TTTKGPTSTSDRPA-- 769
Db 2581 NTKIPCKSPPELTDATSTKRCB--KTRPRKVEKEELSAVERLQTSQGSTHTHKEPASG 2639
QY 770 -----PTTPKE-----TAP 778
Db 2640 DEGIKVLQKAKKPNVEEBSRRRPRAPKKAQPLEDLAGFTLSETSGHTQESITAG 2699
QY 779 TTPKEPATTPKAPATTPPTPTTSEVSTPTTKPTTHKSPDBSTELSAEPTP-- 836
Db 2700 KATKIPCESPLEVVDTTASTKRLHRTVQKVQKEEPSAV--KPTQSGTETTDADKEPAG 2758
QY 837 -----KALENSPKSPGVPTT-----KTPAA--TKPEMTTT 864
Db 2759 EDKGIKALKESAKOTPAASVTSRRRPRAPRESAQAIEDLAGFKDPAAGHTEESMT-- 2816
QY 865 AKDXTTERDLTTFE--TTTAAPKMTKERTATTEKTE-----SKIPATTTQVSTTTQ-- 916
Db 2817 -DDXTTKIPCKSPSELDATSSRRRPRTRAQKVEVEELAVLAVGKLTQTSGETTHTDKXP 2875
QY 917 -----DTTPPKITTLKTTILAPKVTITTKTITTEIANKPDEETAKPKDRA----- 961
Db 2876 VEGSKGTAKF-----QPAKRNVDADVDIGSRQRPRAPKEKAQPLEDLASFPQ 2922
QY 962 -----TNSKATTPKPKPTK-----APK-KPT----- 982
Db 2923 ELSQTPGHTTELANGAADSFTSAPKQTPDGGKPLKISRRVLRAPKVPFVGDVSTRDPVK 2982
QY 983 -----STKKPKTMRVSKPT----- 998
Db 2983 SQSKSNTSLPLPKRGGGKDSVGTGYRLRCMPAPEIIEELPASKKQVAPRAKSS 3042
QY 999 TPT---PRKMTSTMPELNPTSRIAEAMLQTTTRPNQ-----TPNSKLVEVNPKSEDAGGAE 1051
Db 3043 EPVIMKSLRTSAKRIEPAEELNSDMKTNBKHKLQDSVPENKGLSLRSLRQDKTEAE 3102
QY 1052 GETHMLL-----RPHVFMVEVTPDMDYLPRVFNQGIINP 1087
Db 3103 QQITEVFVLAERIEINRNEKKP-----MKTSPEND-----IQNP 3136

```

```

RESULT 15
A41819
proline-rich peptides 637K precursor, prostatic - rat
C:Species: Rattus norvegicus (Norway rat)
C>Date: 13-May-1992 #sequence revision 13-May-1992 #text_change 09-Jul-2004
C:Accession: A42663; A41819; A31966; B20593; A20593
R:De Clercq, N.; Hemschoote, K.; Devos, A.; Peeters, B.; Heyns, W.; Rombauts, W.
J. Biol. Chem. 267, 9884-9894, 1992
A:Title: The 4.4-kilodalton proline-rich polypeptides of the rat ventral prostate are the
le exon.
A:Reference number: A42663; MUID:92250652; PMID:1577819
A:Accession: A42663
A:Molecule type: DNA; mRNA
A:Residues: 1-3706; I, 3708-4077, F, 4079-4155, S, 4157-5761 <DE2>
A:Cross-references: UNIPROT:Q63455; GB:M86514
A:Experimental source: Ventral prostate
A>Note: sequence inconsistent with the nucleotide translation
A>Note: sequence extracted from NCBI backbone (NCBI:100347, NCBI:100348)
R:De Clercq, N.; Hemschoote, K.; Devos, A.; Peeters, B.; Heyns, W.; Rombauts, W.
submitted to the Protein Sequence Database, April 1992
A:Reference number: A41819
A:Accession: A41819
A:Status: not compared with conceptual translation
A:Molecule type: DNA
A:Residues: 1-5762 <DE1>
R:Hemschoote, K.; Peeters, B.; Dirckx, L.; Claessens, F.; De Clercq, N.; Heyns, W.; Wind
J. Biol. Chem. 263, 19159-19165, 1988
A:Title: A single 12.5-kilobase androgen-regulated mRNA encoding multiple proline-rich po
A:Reference number: A31966; MUID:89066721; PMID:3198617
A:Accession: A31966
A:Molecule type: mRNA
A:Residues: 3372-3540 <HEM>
A:Cross-references: GB:M20721; GB:J04188; NID:G206397; PIDN:AAA41950.1; PID:G554494
R:Peeters, B.; Heyns, W.; Bossyns, D.; Rombauts, W.
J. Biol. Chem. 258, 14206-14211, 1983
A:Title: Proline-rich polypeptides bound to rat prostatic binding protein. The primary st
A:Reference number: A94675; MUID:84061859; PMID:6685733
A:Accession: B20593
A:Molecule type: protein
A:Residues: 2020-2057 <PE>
A>Note: this peptide, designated proline-rich polypeptide V, can be found at several loc
A:Molecule type: protein
A:Residues: 2822-2859 <PE2>
A>Note: this peptide, designated proline-rich polypeptide IV, can be found at several loc
C:Genetics:
A:Introns: status absent
A>Note: single copy gene with no introns
C:Superfamily: rat prostatic proline-rich peptides 637K precursor
C:Keywords: prostate; tandem repeat
F:1-26/Domain: signal sequence #status predicted <SIG>
F:27-5761/Product: prostatic proline-rich peptides 637K proprotein #status predicted <MAI
Query Match 14.3%; Score 833.4; DB 2; Length 5762;
Best Local Similarity 13.4%; Pred. No. 1e-08;
Matches 428; Conservative 144; Mismatches 4.9; Indels 2196; Gaps 115;
QY 4 KTLPIYLLLLSVFVIQVSSQDLSSCAGRCGEGYSRDATCNDYNCQ----- 51
Db 456 ETLPIHEI-----QPTQNEYGYQLPNVTVRPVDVA 486
QY 52 -----HYMECPDFKRVCTAELSCKRCFSEFGRZ----- 83
Db 487 LTVTSEPKYETESFLAPQEFVFALEYSND-----VEFFVNEEPPVQAP 531
QY 84 -----CCDCAQCKYDK----- 95
Db 532 ETPGESQFESQLVEPQAQATEVDEBFKTSATFEQLAQFPENDEVTVLPNHYQAQHSILS 591
QY 96 -----CCPYESFCAE----- 106
Db 592 NVTQDQLDITITEKPMEMGTSPVYDAAAPAEVEFLSDQQGVLSQSLEPIYLSLQ 651

```


QY 107 -----EHSYSEN----- 113
Db 652 OBYTTGISQISEGGEFFTOQETPEHVSVMTEEVAQPPGHEVHTGPPGLGHGVHSPALQ 711
QY 114 -----QSSSSSSSSSSSSSTIW----- 130
Db 712 NILTOYSTFPEKEQHSVGLGVPGHLEBFSVEPSPSQENSAHVSVDMLSPVLDLQIF 771
QY 131 ----KTKSSKNSANRELQKLVKDNKNRKKKTKPPFV-----VDEA 172
Db 772 RSTQSKSYKTIKHEDALTI-----TPEFSLDGSILFFQBDLLOPIDST 817
QY 173 GSG-----LDNGDFKVTTPDTS-----TTOHKNKSTSP----- 200
Db 818 GOGERSHIKSTLSLXPPYVSNKSSAQETWSETTYNSKQVDSLSTHLKTSSELPPDYTMG 877
QY 201 ----KITTKAP-----RPSL-----PNSDTSKETS----- 225
Db 878 LEPSLYQOITQSSPSKSHENPKSFFVLKPSAAQALKPRKEXLSPTNNMIPHPSPKPLKN 937
QY 226 ----LTYNKETTIVETKETT-----TINKQ 245
Db 938 MYTHLPAHKNTVPRQIOBQGEYTTISSNGSFQPLDLEVTLSGIIPEVKNHILPKETVNPQ 997
QY 246 TSDGKEKTTSAKETO--SIEKISAKDLAPTSKVLAKTPK-----AETTK--GPAL 294
Db 998 TYSQVKISHSQHVETQHPNSETTVQPLDLBFAINLQPTPKENFAQTQDITTMIGP-- 1055
QY 295 TTPKE-----PPTTPK 306
Db 1056 --PKEVIAQPEHHEGTIPIQDOAEYSTLTVSFQPLDQELTITSEAIRPEPHTVPQ 1113
QY 307 EPASTTPKE-----PTPT-----TIKSAPTTPEPAPTTPKTSAPT 341
Db 1114 QIIIVHPPEPLVHSEQVYVQHNPTEALIQPLDLELTIPTQTAEGE-LPQTLQDST 1172
QY 342 TPKEPAPTTPKEPAP-----TTPKE-----PAPT-----TTKE----- 369
Db 1173 QIIIEP-PTVVVGPIVEEYVQTTSDQAEYPPSPTVSFQSLDLELTISPATRESVHP 1231
QY 370 -----PAPTTKSAPTTPK- 383
Db 1232 SLLQOITVNPPEHPLVHSEQVHTQHPNLTEATVQPLDPLTIPTQPTTEGELPQTLQ 1291
QY 384 ----EPAPTTPKPAPTTP-----KEP- 401
Db 1292 STSQIIIEP-----PIKWALPVVYQVSDQAEYTTSTVSFQPLDQELTITSEAIRPEH 1347
QY 402 APTTPKE-----PPTP-----TPKEPA-----PTTPKEP- 424
Db 1348 HPTVQOITVHTPKHPLVHSEQVHTQHPNTEVTVQPLDLELTMPTQTAEGELPQTLQ 1407
QY 425 APTTPKEPAPTAPKPAPTTPKEPAPTTPKE-----PAPT-----TKEPSPT- 467
Db 1408 STQIIIEP-FTVVVGPIVEEYVQTTSDQAEYPPSPTVSFQSLDGLTITPEPTEHF 1466
QY 468 -----TPKE-----PAPT-----TTKSAP 481
Db 1467 ITOKTIVFPMPYDVTLPQVSVQHLKPTBEGIVQPLDLELTIPTQPTPEGELSQTQEST 1526
QY 482 TTKKEP-----APTTPKSAPTTP-----KEPSP-----TTTKEPA----- 511
Db 1527 TQNKPEHVVAVPVVYCAVTVTPSQYQAEYQKSLQPLDLELTVITSEPTKEAVHSTISK 1586
QY 512 -----PTT-----PKE 517
Db 1587 NSLAINQYVHIQHPNPAEATVQPLDLELTATISSLOQTARGELLYSMQETVTOISEPPKQ 1646
QY 518 -----PAPTTPKPAAP-----TTPKEPAPT 537
Db 1647 VVTPVPEYQEVAVPAPVQDQAKYPLSSIVLSNLSLDQELTILSSBELLGEAHLTTPDETVMVL 1706

QY 538 TPK-----EPAP 544
Db 1707 PPXDRQGIYPDHDHKKHLNLTVEVTPHFLEHTVQHQPTIEERSQSIOKKTQITTEPG- 1765
QY 545 TTTTKKAPAPKPAPTTP-----KETAPTTP-----KKLT----- 575
Db 1766 ----KKVFLAQSSEEVITIPMILKETAPPTPHSMALQSLDEKLTIIHSHSPGWTQOAHANLK 1822
QY 576 ----- 575
Db 1823 ESKGHTTGKILLDYABPNMBIELKHGHLFLTKTEATTESENTNOMTKSLKQVTAFTQN 1882
QY 576 -----PTTPEKL----- 582
Db 1883 KKSMLPALVESQDESOPPPNMSLOPLDQELTSSQSHGVPHIPNTEPKIYLHYABPTG 1942
QY 583 ----APTTPKPAPTT-----PEELAPTPEE-----PTTTPPEBPAP- 616
Db 1943 PFVEPPPELFLKTTKSRFVQGTATQMAASPKEVMVSRAPEKAEVLSGFGDQDESPPPN 2002
QY 617 -----TTPKAAAPNTPK-----EPAPTTP----- 635
Db 2003 MSLQSLDQELTSSQSHGVPHIPNTHGKIYLHYABP-PTGPFVEPPDLFFLTKTTSKPV 2061
QY 636 ----KEPAPTTPK----- 644
Db 2062 EWTLTRIDKSRKEMVSSQPKYEAVALPVHGEQESRSPENMSLQSLQELTILSSQPHGW 2121
QY 645 ----EPAPTTP-----KETAPTTPKGTAPTTLKEP----- 670
Db 2122 VPHPNTHGKIYLHYABP-PTGPFVEPPDLFFLRTTKSKPVQGTATRMVKSPEEMVSLDP 2180
QY 671 -----APTTPKPK----- 678
Db 2181 ENKEAVFPAQGEKGESPPSNMSLQSLDHELFMSSQPHGWIHPPTKPKIYLHYABEP 2240
QY 679 ----APKELAPTTPK----- 689
Db 2241 TGPVFEPPDLFLTTTKSKPVQGTTEMAXSPKEMVQOTPEYKEAVLSGFGDQDESPP 2300
QY 690 ----EPT-----STTSDK 698
Db 2301 PNTSLKSLDQEVAVSSQPHGVPHPPKTPGKIYLHSTIEPPPPPVKPTDLILVKTITKSK 2360
QY 699 PAPTPK-----GTAPTTPKBP----- 716
Db 2361 PAEWTPRDIRDKLKMVPHSPYEBAVFPAGHEQDESPPNMPLOPLDQELTSSQPH 2420
QY 717 ----PPTP-----KEPAPTTPKGTAP-----TTLKEPAPTTPKPPA--PKELA-- 753
Db 2421 GWVPHHPNTPGKIYLHYABP-PTGPFVEPPDLFFLRTTKSKPVQGSPEIAKSPKEMVSQ 2479
QY 754 -----PTTTKG----- 759
Db 2480 TPEYKEAVLSGFGDQDESPPSNMSLQSLDQEVMTSSQPHSGVPHPPKTPGKIYLHSTIE 2539
QY 760 ----PT-----STTSDKPAPTTP-----KETAPTTPKE-----PA----- 785
Db 2540 PPPGPFVKTDLILVKTITKSPAEWTPRRIDKLLKEMVPHSPYEBAVFPAGHEQDES 2599
QY 786 ----PPTPKP-----APTTPETP-----PPTTSEVSTP-----TTTK 814
Db 2600 GSPPNMPLOPLDQELTSSQSHGVPHHPNTPGKIYLHYABPTGPFVEPPDLFLRTPK 2659
QY 815 E-----PTTIHKSPE-----STPELS----- 831
Db 2660 SKFVQGTITQMAKSPPEMVSLSPKNKETVFPAGKQDESPPNMSLQSLDQELTSSQ 2719
QY 832 ----AEP-----TPKALENSPK----- 845
Db 2720 PHGWPHHPNTHGKIYLHYABPTGPFVEPPDLFFLTKTKSKPMQSPQIDKSPKEMFT 2779
QY 846 ----PGVP----- 849

```
Db 2780 QSPYEESLLPAHAEGQESRAPPHESLQPLDQELSLSSHPHGWIHPHNTDPDKIYLHYA 2839
Qy 850 -----TTKTPAATKP-EMTTAKDKTTERDLRTTPTTTAA----- 884
Db 2840 EPPTGPFVEPPDLFFLTKTK-----SKPVQGTATKTDKSPEDRVSTQPYEKEAVLSGPGED 2895
Qy 885 -----PKMT-----KETATITE----- 896
Db 2896 QDESPPPNMSLQSLDQELAISSQPHGWIHPHNPADKILYHYABPTGPFVEPPDLFFL 2955
Qy 897 KITESK-ITATTTQVTSITTQ----- 916
Db 2956 KITKSKPLQGTPTQMAKSPKEMVSTQPYEKEADLSAPGENRDESSPPNNMSLHPLDQELS 3015
Qy 917 -----DTTP-----KITTLKTTTLAPKVTTTKTITTTE 946
Db 3016 LSSQPHGWIHPHNTHGKIYLHYABPTGPFVEPPDLFFLTKTKSP----- 3062
Qy 947 INMKPEETAK-PKD----- 959
Db 3063 VQGSFQIDKSPKEVFSQSESESVLPAQAEQESRAPPHMSLQLLDQKLTLSHPHG 3122
Qy 960 -----RATNSKATTPKPKTKAPK----- 980
Db 3123 WIPHPNTGKIYLHYABPTGPFVEPPDLFFLTKTKGKPVQGSPRQVDKSPKEMFTQSP 3182
Qy 981 -----PTSTKK-----PK 988
Db 3183 EYESVLPAAEQEESRAPPHMSLQPLDQDLTSSHPHGWIHPHNTDPDKIYLHYABPP 3242
Qy 989 TMRVRKRP-----KTTP-----PRKMTSTMP-----LNP 1014
Db 3243 TGPVVEPPDLFFLRATKSKPAQWPTQMAKSPPEMVSLSPENKENVFPAAQKGQESISIP 3302
Qy 1015 TSRIAEAMLQTT-----RPN-----QTPNSKL 1037
Db 3303 PHMSLQPLDQDLTPSSHPHGWIHPHNTHGKIYLHYABPTGPFVEPPDLFFLTKTKSP 3362
Qy 1038 VEVNP----- 1042
Db 3363 VQVSPRQIDKSPKEVFTQSPYEESVLPAAEQEESRAPPHMSLQPLDQDLTSSHPHG 3422
Qy 1043 -----KSEDAG----- 1049
Db 3423 WITHHPNTHGKIYLHYABPTGPFVEPPDLFFLTKTKSPVHGSFRQIDKSHKEMFTQSP 3482
Qy 1050 -----AEGE-----TPHMLLR-----PHVEMPE--VTPDMDYLPRV-PN 1080
Db 3483 EYESVLPAAEQEESRAPPHMSLQPLDQDLTSSHPHGWIHPHNTDPDKIYLHYABPP 3542
Qy 1081 QGIINP 1087
Db 3543 TGPVVEP 3549
```

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: October 13, 2004, 11:23:49 ; Search time 88.5238 Seconds
(without alignments)
4449.477 Million cell updates/sec

Title: SEQ1-C
Perfect score: 5887
Sequence: 1 MAWKTLPIYLLLSVFVQI.....DMDYLPVPNGQIINPMLS 1098

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.1

Searched: 2002273 seqs, 358729299 residues

Total number of hits satisfying chosen parameters: 2002273

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A_Geneseq_238sep04:.*
1: Geneseqp1980s:.*
2: Geneseqp1990s:.*
3: Geneseqp2000s:.*
4: Geneseqp2001s:.*
5: Geneseqp2002s:.*
6: Geneseqp2003as:.*
7: Geneseqp2003bs:.*
8: Geneseqp2004s:.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | ID | Description |
|------------|--------|-------------|--------|------------|-------------|
| 1 | 5872.8 | 99.8 | 1299 | 4 AAM24322 | Human EST |
| 2 | 5872.8 | 99.8 | 1404 | 2 AAB26049 | MSF precu |
| 3 | 5872.8 | 99.8 | 1404 | 4 AAB29773 | Human meg |
| 4 | 5872.8 | 99.8 | 1404 | 4 AAB60568 | Human meg |
| 5 | 5872.8 | 99.8 | 1404 | 8 ADM98014 | Human meg |
| 6 | 5872.8 | 99.8 | 1415 | 4 AAU32262 | Novel hum |
| 7 | 5863.8 | 99.6 | 1404 | 7 ADK65839 | Angiogene |
| 8 | 5807.9 | 95.3 | 1311 | 8 ADK67912 | Human ext |
| 9 | 5598.8 | 95.1 | 1320 | 7 ADK65819 | Angiogene |
| 10 | 5343.8 | 90.8 | 1270 | 8 ADK67911 | Human ext |
| 11 | 2920 | 49.6 | 546 | 4 ABUS3252 | Human tes |
| 12 | 2888 | 49.1 | 551 | 4 ABUS3253 | Human tes |
| 13 | 2880.1 | 48.9 | 902 | 4 AAB29778 | Human MSF |
| 14 | 2757 | 46.8 | 513 | 4 ABUS3254 | Human tes |
| 15 | 2222.8 | 37.8 | 452 | 2 AAR80041 | Human meg |
| 16 | 1866.8 | 26.6 | 232 | 5 AAU11261 | Human HAP |
| 17 | 1409.9 | 23.9 | 538 | 5 AAO18834 | 3' cartil |
| 18 | 1396.9 | 23.7 | 5179 | 4 AAM24516 | C899P pre |
| 19 | 1396.9 | 23.7 | 5179 | 6 ABP55365 | Human col |
| 20 | 1396.9 | 23.7 | 5179 | 6 ABO07258 | Human p53 |
| 21 | 1396.9 | 23.7 | 5179 | 7 AD448091 | Human Pro |
| 22 | 1396.9 | 23.7 | 5179 | 7 AD448091 | Human Pro |
| 23 | 1132 | 19.2 | 1664 | 2 AAW43106 | C. thermo |
| 24 | 1127.3 | 19.1 | 8991 | 4 ABU08487 | S. pneumo |
| 25 | 1100.7 | 18.7 | 1795 | 4 ABB69806 | Drosophil |

| | | | | | |
|----|--------|------|------|------------|-----------|
| 26 | 1100.4 | 18.7 | 717 | 4 ABUS3144 | Human tes |
| 27 | 1097 | 18.6 | 214 | 4 ABUS3255 | Human tes |
| 28 | 1067.7 | 18.1 | 763 | 3 AAG38942 | Arabidops |
| 29 | 1048.4 | 17.8 | 1049 | 4 ABB61364 | Drosophil |
| 30 | 1047.2 | 17.8 | 2284 | 4 ABB71434 | Drosophil |
| 31 | 1032.6 | 17.5 | 5703 | 8 ADL23265 | Human MUC |
| 32 | 1009.9 | 17.2 | 188 | 5 AAO18833 | 5' cartil |
| 33 | 995.3 | 16.9 | 4315 | 5 ABP43908 | MUC5B par |
| 34 | 987.9 | 16.8 | 770 | 4 ABUS3141 | Human tes |
| 35 | 981 | 16.7 | 778 | 4 ABUS3143 | Human tes |
| 36 | 975 | 16.6 | 2768 | 4 ABB68397 | Drosophil |
| 37 | 974.4 | 16.6 | 692 | 4 ABUS3155 | Human tes |
| 38 | 968.2 | 16.4 | 1460 | 7 ADN39110 | Cancer/an |
| 39 | 967.7 | 16.4 | 1538 | 7 ADI21202 | Novel hum |
| 40 | 965.9 | 16.4 | 2112 | 4 ABB60403 | Drosophil |
| 41 | 964.1 | 16.4 | 717 | 4 ABUS3150 | Human tes |
| 42 | 964.1 | 16.4 | 717 | 4 ABUS3149 | Human tes |
| 43 | 964.1 | 16.4 | 717 | 4 ABUS3151 | Human tes |
| 44 | 964.1 | 16.4 | 717 | 4 ABUS3145 | Human tes |
| 45 | 964.1 | 16.4 | 717 | 4 ABUS3148 | Human tes |

ALIGNMENTS

RESULT 1
AAM24322
ID AAM24322 standard; protein; 1299 AA.
XX
AC AAM24322;
XX
DT 12-OCT-2001 (first entry)
XX
DE Human EST encoded protein SEQ ID NO: 1847.
XX
KW Human, sheep; pig; cow; fruit fly; yeast; hamster; macaque; horse;
KW tomato; monkey; dog; sea urchin; expressed sequence tag; EST;
KW diagnostics; forensic test; gene mapping; genetic disorder; biodiversity;
KW gene therapy; nutrition.
XX
OS Homo sapiens.
XX
PN WO200154477-A2.
XX
PD 02-AUG-2001.
XX
PF 25-JAN-2001; 2001WO-US002687.
XX
PR 25-JAN-2000; 2000US-00491404.
PR 17-JUL-2000; 2000US-00617746.
PR 03-AUG-2000; 2000US-00631451.
PR 15-SEP-2000; 2000US-00663870.
XX
PA (HYSE-) HYSEQ INC.
PI Tang YT, Liu C, Zhou P, Qian XB, Wang Z, Chen R, Asundi V;
PI Cao Y, Drmanac RA, Zhang J, Werhman T;
DR WPI: 2001-476164/51.
DR N-PSDB: AAH98981.
XX
PT Isolated polypeptide for treatment of diseases, diagnostics, raising
PT antibodies and research use.
XX
PS Claim 20; Page 1198-1201; 1275pp; English.
XX
CC The present invention provides the protein and coding sequences of novel
CC proteins from a variety of organisms, including human, dog, cat, horse,
CC cow, pig, hamster, monkey, macaque, yeast, bacteria, fruit fly, sea
CC urchin and tomato. These were derived from expressed sequence tags (ESTs)
CC from the organism of interest. They can be used in diagnostics,
CC forensics, gene mapping, identification of mutations, to assess
CC biodiversity and for nutritional purposes. The present sequence is a

XX Claim 1, 2 and 3; Fig 1; 87pp; English.

XX The sequence given is a full length translation from the megakaryocyte

CC stimulating factor (MSF) precursor. The sequence covered by exons II, III

CC and IV encodes megakaryocyte stimulating factor (MSF). This sequence is

CC modified by the addition of an N-terminal sequence encoding a secretory

CC leader, an initiating methionine preceding exon II and a terminating

CC codon following exon IV. The cDNA sequence given contains sequences

CC derived from human megakaryocyte colony stimulating factor (meg-CSF).

CC Exon I contains the initiating methionine, and encodes a classical

CC mammalian protein secretion signal sequence. The sequence encoding the

CC original meg-CSF includes exons II-IV and is thought to terminate in the

CC region between amino acid residues 134 - 147. The primary transcript of

CC this gene may be cleaved in different ways to yield a family of mRNAs

CC each encoding a different MSF protein. Exons V and VI are thought to be

CC related to the activity of the factor and are also implicated in the

CC stability, folding and processing of the molecule. These exons are also

CC thought to play a role in the observed synergy of MSF with other

CC cytokines. Exons V - XII are believed to be implicated in the processing

CC or folding of the appropriate structure of the resulting factor, i.e. one

CC or more of these exons may contain sequences which direct proteolytic

CC cleavage, adhesion, organisation of the cellular matrix or extracellular

CC matrix processing. Both naturally occurring and non-naturally occurring

CC MSF's may be characterised by various combinations of alternatively

CC spliced exons from this sequence, with the exons spliced together in

CC differing orders to form different members of the MSF family. (Updated on

CC 25-MAR-2003 to correct PN field.)

XX SQ Sequence 1404 AA;

Query Match 99.8%; Score 5872.8; DB 2; Length 1404;

Best Local Similarity 96.3%; Pred. No. 3.2e-154;

Matches 1098; Conservative 0; Mismatches 0; Indels 42; Gaps 1;

QY 1 MAWKTLPIYLLLLSVFIQVSSODLSCACGCGEGSRDATCNCYNCOHYMECCPDF 60

DB 1 MAWKTLPIYLLLLSVFIQVSSODLSCACGCGEGSRDATCNCYNCOHYMECCPDF 60

QY 61 KRVTAEALSCGRCFESPERGECDCDAQCKYDCCPDYSCFAEVHNPTSPSSKKAP 120

DB 61 KRVTAEALSCGRCFESPERGECDCDAQCKYDCCPDYSCFAEVHNPTSPSSKKAP 120

QY 121 PPSGASQIKSTKTSPPNKKTKVIESEITE----- 156

DB 121 PPSGASQIKSTKTSPPNKKTKVIESEITE----- 180

QY 157 -----KVQDKNNRTKKKTPKPPVVDVDEAGSLDNGDFKVTTFDTST 198

DB 181 KTSSGNSAANRELQKLVKQDKNNRTKKKTPKPPVVDVDEAGSLDNGDFKVTTFDTST 240

QY 199 TQHNKVSTSPKITTAKPINRPSLPNSDTSKETSILVNKETTETKTTNKQSTSDG 258

DB 241 TQHNKVSTSPKITTAKPINRPSLPNSDTSKETSILVNKETTETKTTNKQSTSDG 300

QY 259 KEKTTSAKETQSIKETSADLAPTSKVLAKPTPKAETTTKGPALTTPKPTTTPKEPAS 318

DB 301 KEKTTSAKETQSIKETSADLAPTSKVLAKPTPKAETTTKGPALTTPKPTTTPKEPAS 360

QY 319 TTPKEPTPTTIKSAPTTPKAPATTTKSAPTTTPKBPATTTKBPATTTKBPATTTKBP 378

DB 361 TTPKEPTPTTIKSAPTTPKAPATTTKSAPTTTPKBPATTTKBPATTTKBPATTTKBP 420

QY 379 APTTKSAPTTPKAPATTTPKAPATTTPKBPATTTKBPATTTKBPATTTKBPATTTKBP 438

DB 421 APTTKSAPTTPKAPATTTPKAPATTTPKBPATTTKBPATTTKBPATTTKBPATTTKBP 480

QY 439 EPAPTAKPKPAPTTPKAPATTTPKAPATTTPKBPATTTKBPATTTKBPATTTKBPATTT 498

DB 481 EPAPTAKPKPAPTTPKAPATTTPKAPATTTPKBPATTTKBPATTTKBPATTTKBPATTT 540

QY 499 TTKSAPTTKBPSPPTTKBPATTTPKBPATTTPKBPATTTPKBPATTTPKBPATTTKBP 558

DB 541 TTKSAPTTKBPSPPTTKBPATTTPKBPATTTPKBPATTTPKBPATTTPKBPATTTKBP 600

QY 559 APTAPKEPAPTTPKETAPTTPKLTPTTPEKLAAPTTPEKAPATTPEKAPATTPEEETPTT 618

DB 601 APTAPKEPAPTTPKETAPTTPKLTPTTPEKLAAPTTPEKAPATTPEKAPATTPEEETPTT 660

QY 619 PEEPAATTPKAAAPNTPEKAPATTPEKAPATTPEKAPATTPEKAPATTPEKAPATTPEE 678

DB 661 PEEPAATTPKAAAPNTPEKAPATTPEKAPATTPEKAPATTPEKAPATTPEKAPATTPEE 720

QY 679 APTTPKKPAPKELAPTTKEPTSTTSKAPATTPKGTAPTTPEKAPATTPEKAPATTPEK 738

DB 721 APTTPKKPAPKELAPTTKEPTSTTSKAPATTPKGTAPTTPEKAPATTPEKAPATTPEK 780

QY 739 TAPTTLKEPAPTTPKKAPKELAPTTTKGPTSTTSKAPATTPKGTAPTTPEKAPATTPEK 798

DB 781 TAPTTLKEPAPTTPKKAPKELAPTTTKGPTSTTSKAPATTPKGTAPTTPEKAPATTPEK 840

QY 799 KPAPTTPETPPPTSEVSTPTTKETTTTKSDESTPELSAAPTPEKALENSKEPGVPT 858

DB 841 KPAPTTPETPPPTSEVSTPTTKETTTTKSDESTPELSAAPTPEKALENSKEPGVPT 900

QY 859 TKTPAATKPEMTTAKDKTTERDLRTPETTTAAPKMTKETATTTTEKTTESKITATTQV 918

DB 901 TKTPAATKPEMTTAKDKTTERDLRTPETTTAAPKMTKETATTTTEKTTESKITATTQV 960

QY 919 TSTTTQDTPFKITLTKTTTLAPKVTTKKIITTTIMNKPBEETAKPKDRATNSKATTPK 978

DB 961 TSTTTQDTPFKITLTKTTTLAPKVTTKKIITTTIMNKPBEETAKPKDRATNSKATTPK 1020

QY 979 PQKPTKAPKPTSTKXPKTMPRVKPKTTPTRKMTSTMPELNPTSRIAEAMLOTTTRPN 1038

DB 1021 PQKPTKAPKPTSTKXPKTMPRVKPKTTPTRKMTSTMPELNPTSRIAEAMLOTTTRPN 1080

QY 1039 QTNSKLIVEVNPKSEDAGAGETPHMLLRPHVFMPEVTPDMDYLPRVNPQGIINPMLS 1098

DB 1081 QTNSKLIVEVNPKSEDAGAGETPHMLLRPHVFMPEVTPDMDYLPRVNPQGIINPMLS 1140

RESULT 3

AAB29773

ID AAB29773 standard; protein; 1404 AA.

XX AAB29773;

XX AC AAB29773;

DT 28-FEB-2001 (first entry)

XX Human megakaryocyte stimulating factor (MSF), SEQ ID NO:1.

DE Human MSF; megakaryocyte stimulating factor; tribonectin;

KW alternative splicing; joint boundary lubricant; O-linked oligosaccharide;

KW osteoarthritis; tribosupplementation; tissue adhesion inhibition;

KW friction coefficient reduction; gene therapy; antiarthritic; osteopathic.

XX Homo sapiens.

OS WO200064930-A2.

XX PN 02-NOV-2000.

XX PD 24-APR-2000; 2000WO-US010953.

XX PF 23-APR-1999; 99US-00298970.

XX PR (RHOD-) RHODE ISLAND HOSPITAL LIFESPAN PARTNER.

PA Jay GD;

XX WPI; 2001-024673/03.

XX DR N-PSDB; AAC81498.

XX Novel tribonectin polypeptide useful as lubricant for treating

PT osteoarthritis, comprises O-linked lubricating moiety.

XX PS Claim 3; Page 7; 47pp; English.

CC The invention relates to a human tribonectin which is a product of

CC alternative splicing of the human MSF (megakaryocyte stimulating factor)

CC gene. The tribonectin has at least one O-linked oligosaccharide

CC lubricating moiety and has a polypeptide sequence comprising 1-76 repeats

CC of a motif having at least 50% identity to the sequence KEPATT

CC (AAB25774). The invention also relates to a nucleic acid encoding a human

CC MSF-derived tribonectin; a biocompatible composition comprising a human

CC tribonectin for inhibiting tissue adhesion formation; and a method of

CC diagnosing osteoarthritis or a predisposition to osteoarthritis by

CC measuring the amount of MSF or its fragment in a biological sample of a

CC mammal, wherein an increased amount of MSF compared to a control

CC indicates the presence of or predisposition to developing osteoarthritis.

CC The tribonectin and DNA encoding it are useful in the treatment of

CC osteoarthritis, where they may be used for lubricating mammalian joints,

CC such as articulating joints of humans, dogs or horses. The tribonectin,

CC when formulated as a membrane, foam, gel or fibre, is useful for

CC inhibiting adhesion between two surfaces such as the injured tissues of a

CC mammal, where the injury is caused by a surgical insertion or trauma, or

CC an artificial device e.g., an orthopaedic implant. In particular, one of

CC the surfaces is pericardial tissue. DNA encoding a tribonectin may be

CC used in gene therapy. The present sequence represents human MSF

XX Sequence 1404 AA;

Query Match 99.8%; Score 5872.8; DB 4; Length 1404;

Best Local Similarity 96.3%; Pred. No. 3.2e-154; Mismatches 0; Indels 42; Gaps 1;

Matches 1098; Conservative 0;

QY 1 MAWKTLPIYLILLVSVFVIQVSSODLSACAGROGEGYSRDATCNCDCVNCQHYMECCPDF 60

DB 1 MAWKTLPIYLILLVSVFVIQVSSODLSACAGROGEGYSRDATCNCDCVNCQHYMECCPDF 60

QY 61 KVVCTAELSCGRCPESFERGECDCDAQCKYDKCCPDYDFCAEVHNPTSPSSKKAP 120

DB 61 KVVCTAELSCGRCPESFERGECDCDAQCKYDKCCPDYDFCAEVHNPTSPSSKKAP 120

QY 121 PPSGASQIKSTTKSPKPPNKKTKVIESEITE----- 156

DB 121 PPSGASQIKSTTKSPKPPNKKTKVIESEITEHSVSENQESSSSSSSSSSSTI 180

QY 157 -----KVQDNKNRTKKPKPPVVDAGSLGNGDFKVTPTDST 198

DB 181 KIKSSKSAANRELQKLVKDNKNRTKKPKPPVVDAGSLGNGDFKVTPTDST 240

QY 199 TQHNKVSTSPKITTAKPINRPSLPNNSDTSKSLTVNKETTVETKTTNKTQSTDG 258

DB 241 TQHNKVSTSPKITTAKPINRPSLPNNSDTSKSLTVNKETTVETKTTNKTQSTDG 300

QY 259 KEKTSKETSQIKTSKADLAFTSKVLAKPTPKAETTTKGPALTTKPEPTTPKBPAS 318

DB 301 KEKTSKETSQIKTSKADLAFTSKVLAKPTPKAETTTKGPALTTKPEPTTPKBPAS 360

QY 319 TTPKEPTTTIKSAPTTKBPATTPKBPATTPKBPATTPKBPATTPKBPATTPKBP 378

DB 361 TTPKEPTTTIKSAPTTKBPATTPKBPATTPKBPATTPKBPATTPKBPATTPKBP 420

QY 379 APTTTKSAPTTKBPATTPKBPATTPKBPATTPKBPATTPKBPATTPKBPATTPK 438

DB 421 APTTTKSAPTTKBPATTPKBPATTPKBPATTPKBPATTPKBPATTPKBPATTPK 480

QY 439 EPAPTAPKBPATTPKBPATTPKBPATTPKBPATTPKBPATTPKBPATTPKBPAT 498

DB 481 EPAPTAPKBPATTPKBPATTPKBPATTPKBPATTPKBPATTPKBPATTPKBPAT 540

QY 499 TTKSAPTTKBPSTTTKBPATTPKBPATTPKBPATTPKBPATTPKBPATTPKBPAT 558

DB 541 TTKSAPTTKBPSTTTKBPATTPKBPATTPKBPATTPKBPATTPKBPATTPKBPAT 600

QY 559 APTAPKBPATTPKBPATTPKBPATTPKBPATTPKBPATTPKBPATTPKBPATTP 618

DB 601 APTAPKBPATTPKBPATTPKBPATTPKBPATTPKBPATTPKBPATTPKBPATTP 660

QY 619 PEPAPATTTKAAAPNTPKBPATTPKBPATTPKBPATTPKBPATTPKBPATTPKBPAT 678

DB 661 PEPAPATTTKAAAPNTPKBPATTPKBPATTPKBPATTPKBPATTPKBPATTPKBPAT 720

QY 679 APTTPKBPAPKELAPTTTKEPTSTTSKBPATTPKBPATTPKBPATTPKBPATTPK 738

DB 721 APTTPKBPAPKELAPTTTKEPTSTTSKBPATTPKBPATTPKBPATTPKBPATTPK 780

QY 739 TAPTTLKBPATTPKBPAPKELAPTTTKEPTSTTSKBPATTPKBPATTPKBPATTPK 798

DB 781 TAPTTLKBPATTPKBPAPKELAPTTTKEPTSTTSKBPATTPKBPATTPKBPATTPK 840

QY 799 KPAPTTPEPTPTTSEVSTPTTKEPTTIHKSPDSTPELSAEPKPALENSPKFPGVPT 858

DB 841 KPAPTTPEPTPTTSEVSTPTTKEPTTIHKSPDSTPELSAEPKPALENSPKFPGVPT 900

QY 859 TKTPAATKEMTTAKDKTTERDLRTTPETTTAAAPQWTKETATTTKTESKITATTQV 918

DB 901 TKTPAATKEMTTAKDKTTERDLRTTPETTTAAAPQWTKETATTTKTESKITATTQV 960

QY 919 TSTTTQDTPPFKITTLLKTTTLAPKVTITTKTITTTIMNKPEETAKPKDRATNSKATTPK 978

DB 961 TSTTTQDTPPFKITTLLKTTTLAPKVTITTKTITTTIMNKPEETAKPKDRATNSKATTPK 1020

QY 979 POKPTKAPKPTSTKPKTMPRVKPKTPTPRKMTSTMPKLTSTMPKLTSTMPKLTSTMP 1038

DB 1021 POKPTKAPKPTSTKPKTMPRVKPKTPTPRKMTSTMPKLTSTMPKLTSTMPKLTSTMP 1080

QY 1039 QTPNSKLVEVNPKSEDAGAGGETPHMLLRPHVFWPEVTPDMYDLPVFNQGIINPMLS 1098

DB 1081 QTPNSKLVEVNPKSEDAGAGGETPHMLLRPHVFWPEVTPDMYDLPVFNQGIINPMLS 1140

RESULT 4

AA60568

ID AAB60568 standard; protein; 1404 AA.

XX AC

XX AAB60568;

XX DT 27-APR-2001 (first entry)

XX DE Human megakaryocyte stimulating factor (MSF, CACP).

XX KW Human; CACP protein; camptodactyly-arthropathy-coxa vara-pericarditis;

XX KW MSF; megakaryocyte stimulating factor; synovial lubricant;

XX KW chromosome 1q25-31; osteoarthritis; joint lubrication; osteopathic;

XX KW antiarthritic.

OS Homo sapiens.

XX WO200107068-A1.

PN 01-FEB-2001.

PD 21-JUL-2000; 2000WO-US020002.

PF 23-JUL-1999; 99US-0145328P.

PR 19-JUL-2000; 2000US-00145328.

XX (UYCA-) UNIV CASE WESTERN RESERVE.

XX Warman ML;

XX WPI; 2001-182721/18.

XX New composition comprising the camptodactyly-arthropathy-coxa vara-

XX pericarditis protein in combination with an anesthetic, useful for

XX treating osteoarthritis, or as lubricants of tissue and joints.

XX Example 1; Page; 34pp; English.

CC The invention relates to a method of treating osteoarthritis via the
CC administration of a composition comprising the camptodactyl-arthropathy-
CC coxa vara-pericarditis (CACP) protein, or portions of the CACP protein.
CC The composition may further comprise a local anaesthetic. The composition
CC of the invention may be administered via intra-articular or intravenous
CC injection. The human CACP protein is identified in the invention as being
CC megakaryocyte stimulating factor (MSF). The gene encoding CACP protein
CC (MSF) is located on chromosome 1q25-31, and mutations in this gene are
CC responsible for the heritable disorder camptodactyl-arthropathy-coxa
CC vara-pericarditis, in which patients have synovial hyperplasia without
CC evidence of inflammation. CACP protein (MSF) acts as a synovium
CC lubricant, and can be used to lubricate tissue and joints in the
CC treatment of osteoarthritis. The composition may be applied to reduce the
CC symptoms of osteoarthritis (e.g., joint pain, loss of range of movement
CC or joint damage). The present sequence represents human megakaryocyte
CC stimulating factor (MSF, CACP protein). Note: This sequence is not given
CC in its entirety in figure 4 of the specification, although a GenBank
CC accession number was given. This sequence was therefore obtained from
CC GenBank (U70316)
XX
XX
SQ Sequence 1404 AA;
Query Match 99.8%; Score 5872.8; DB 4; Length 1404;
Best Local Similarity 96.3%; Pred. No. 3.2e-154;
Matches 1098; Conservative 0; Mismatches 0; Indels 42; Gaps 1;
QY 1 MAWKTLPIYLILLVSVFIQVSSODLSACAGCGGYSRDATCNCDYNCOHMECCPDF 60
DB 1 MAWKTLPIYLILLVSVFIQVSSODLSACAGCGGYSRDATCNCDYNCOHMECCPDF 60
QY 61 KEVCTAELSCKRCFESFERGECDCDAQCKYDKCCPDYSEFCAEVHNPTSPSSKKAP 120
DB 61 KEVCTAELSCKRCFESFERGECDCDAQCKYDKCCPDYSEFCAEVHNPTSPSSKKAP 120
QY 121 PPSGASQTIKSTKSPKPNKKTKVIESEITE- 156
DB 121 PPSGASQTIKSTKSPKPNKKTKVIESEITE- 156
QY 157 -----KVKNKKNRTKKKTPPPVVDVDEAGSLDNGDFKVTTPDTST 198
DB 181 KTKSSKNSANRELQKLVKCNKKNRYKKTPPPVVDVDEAGSLDNGDFKVTTPDTST 240
QY 199 TQHNKVSTSPKITTAKPINRPSLPNSDTSKETSLSLVNKTETVTKTTNNKQTSDDG 258
DB 241 TQHNKVSTSPKITTAKPINRPSLPNSDTSKETSLSLVNKTETVTKTTNNKQTSDDG 300
QY 259 KEKTSKETSQIEKTSADLAPTSKVLAKPTPKAETTTKGPALTTKPEPTTPKEPAS 318
DB 301 KEKTSKETSQIEKTSADLAPTSKVLAKPTPKAETTTKGPALTTKPEPTTPKEPAS 360
QY 319 TTPKEPTPTTIKSAPTTKPEAPTTTKSAPTTKPEAPTTTKPEAPTTTKPEAPTTTKPE 378
DB 361 TTPKEPTPTTIKSAPTTKPEAPTTTKSAPTTKPEAPTTTKPEAPTTTKPEAPTTTKPE 420
QY 379 APTTKSAPTTKPEAPTTKPEAPTTKPEAPTTKPEPTTPKPEAPTTKPEAPTTKPE 438
DB 421 APTTKSAPTTKPEAPTTKPEAPTTKPEAPTTKPEPTTPKPEAPTTKPEAPTTKPE 480
QY 439 EPAPAPKPKAPTTKPEAPTTKPEAPTTKPEAPTTKPEAPTTTKSAPTTTKPEAPTT 498
DB 481 EPAPAPKPKAPTTKPEAPTTKPEAPTTKPEAPTTKPEAPTTTKSAPTTTKPEAPTT 540
QY 499 TTKSAPTTKPESPPTTKPEAPTTKPEAPTTKPEAPTTKPEAPTTKPEAPTTTKKP 558
DB 541 TTKSAPTTKPESPPTTKPEAPTTKPEAPTTKPEAPTTKPEAPTTKPEAPTTTKKP 600
QY 559 APTAPKPEAPTTKPEAPTTKPEAPTTKPEAPTTKPEAPTTPEELAPTTPEPTPTT 618
DB 601 APTAPKPEAPTTKPEAPTTKPEAPTTKPEAPTTKPEAPTTPEELAPTTPEPTPTT 660
QY 619 PPEEPAPTTKAAAPNTPKPEAPTTKPEAPTTKPEAPTTKPEAPTTKPEAPTTTKPE 678
DB 661 PPEEPAPTTKAAAPNTPKPEAPTTKPEAPTTKPEAPTTKPEAPTTKPEAPTTTKPE 720

QY 679 APTPKKPKAPKELAPTTTKPEPTSTTSDDKAPPTPKGTASTTKPEAPTTKPEAPTTKPG 738
DB 721 APTTKKPKAPKELAPTTTKPEPTSTTSDDKAPPTPKGTASTTKPEAPTTKPEAPTTKPG 780
QY 739 TAPTTLKPEAPTTPKKAPKELAPTTTKGPTSTTSDDKAPPTPKGTASTTKPEAPTTK 798
DB 781 TAPTTLKPEAPTTPKKAPKELAPTTTKGPTSTTSDDKAPPTPKGTASTTKPEAPTTK 840
QY 799 KPAPTTPEPTTSEVSTPTTKPEPTTHKSPDESTPDSABPTPKALENSKPEGVPT 858
DB 841 KPAPTTPEPTTSEVSTPTTKPEPTTHKSPDESTPDSABPTPKALENSKPEGVPT 900
QY 859 TKTPAAKPKEMTTAKDKTTERDLRTPETTTAAPKMTKETATTTTEKTTESKITATTTOV 918
DB 901 TKTPAAKPKEMTTAKDKTTERDLRTPETTTAAPKMTKETATTTTEKTTESKITATTTOV 960
QY 919 TSTTQDTPFKITTLTKTTTLAPKVTTKKTTITTTIMNKPEETAKPKORATNSKATTPK 978
DB 961 TSTTQDTPFKITTLTKTTTLAPKVTTKKTTITTTIMNKPEETAKPKORATNSKATTPK 1020
QY 979 PQKPTKAPKPTSTKPKTMPRVKPKTTPTPEKMTSTMPELNPTTSRIEAMLOTTTRPN 1038
DB 1021 PQKPTKAPKPTSTKPKTMPRVKPKTTPTPKMTSTMPELNPTTSRIEAMLOTTTRPN 1080
QY 1039 QTNSKLIVEVNPKSSEDAGGAEGETPHMLRPHVFMPEVTPDMDYLPVFNQGIINPMLS 1098
DB 1081 QTNSKLIVEVNPKSSEDAGGAEGETPHMLRPHVFMPEVTPDMDYLPVFNQGIINPMLS 1140
RESULT 5
ADM98014
ID ADM98014 standard; protein; 1404 AA.
XX ADM98014;
AC ADM98014;
XX
DT 01-JUL-2004 (first entry)
XX Human megakaryocyte stimulating factor (MSF).
DE lubricating polypeptide; O-linked oligosaccharide; joint lubrication;
KW CAP; camptodactyl-arthropathy pericarditis; osteoarthritis; human;
KW megakaryocyte stimulating factor; MSF.
XX Homo sapiens.
XX OS
XX US2004072741-A1.
XX 15-APR-2004.
XX 02-JUL-2001; 2001US-008997188.
XX 23-APR-1999; 99US-00298970.
XX 24-APR-2000; 2000US-00556246.
XX (JAYG/) JAY G D.
XX Jay GD;
XX WPI; 2004-373948/35.
XX N-PSDB; ADM98015.
XX New tribonectin polypeptides and polynucleotides for lubricating joints
PT or other tissues to prevent or treat Camptodactyl-arthropathy-
PT pericarditis syndrome or osteoarthritis.
PS Claim 1; SEQ ID NO 1; 34pp; English.
XX The invention relates to a lubricating polypeptide and at least one O-
XX linked oligosaccharide. The composition and methods are useful for
XX lubricating joints or other tissues to prevent or treat camptodactyl-
XX arthropathy pericarditis (CAP) or osteoarthritis in mammals. The present
XX sequence represents the amino acid sequence of the human megakaryocyte

QY 157 -----KVNKKNRRTKKKPTPPVVDVDEAGSLDNGDFKVTPTDST 198
 Db 181 KIKSSKNSAANRELQKLVKVNKKNRRTKKKPTPPVVDVDEAGSLDNGDFKVTPTDST 240
 QY 199 TQHNKVSFSPKITTAKPINRPSLSPNSDTSKETSLSVKNKETTVEKTTTTNNKQTSDDG 258
 Db 241 TQHNKVSFSPKITTAKPINRPSLSPNSDTSKETSLSVKNKETTVEKTTTTNNKQTSDDG 300
 QY 259 KEKTTSAKETQSIKTSKADLAPTQKVLAKPTPKAETTTKGPALTPKPEPTTPKBPAS 318
 Db 301 KEKTTSAKETQSIKTSKADLAPTQKVLAKPTPKAETTTKGPALTPKPEPTTPKBPAS 360
 QY 319 TTPKEPTPTTKSAPTTKPEAPTTKSAPTTKPEAPTTKPEAPTTKPEAPTTKKEP 378
 Db 361 TTPKEPTPTTKSAPTTKPEAPTTKSAPTTKPEAPTTKPEAPTTKPEAPTTKKEP 420
 QY 379 APTTTKSAPTTKPEAPTTKPEAPTTKPEAPTTKPEAPTTKPEAPTTKPEAPTTK 438
 Db 421 APTTTKSAPTTKPEAPTTKPEAPTTKPEAPTTKPEAPTTKPEAPTTKPEAPTTK 480
 QY 439 EPAPTAPKAPPTTKPEAPTTKPEAPTTTKKPSPTTKPEAPTTTKSAPTTTKPEAPT 498
 Db 481 EPAPTAPKAPPTTKPEAPTTKPEAPTTTKKPSPTTKPEAPTTTKSAPTTTKPEAPT 540
 QY 499 TTKSAPTTKPEPSPTTKPEAPTTKPEAPTTKPEAPTTKPEAPTTKPEAPTTTKP 558
 Db 541 TTKSAPTTKPEPSPTTKPEAPTTKPEAPTTKPEAPTTKPEAPTTKPEAPTTTKP 600
 QY 559 APTAPKPEAPTTKPEAPTTKPEAPTTKPEAPTTKPEAPTTKPEAPTTKPEAPTT 618
 Db 601 APTAPKPEAPTTKPEAPTTKPEAPTTKPEAPTTKPEAPTTKPEAPTTKPEAPTT 660
 QY 619 PEAPPTTKAAAPNTKPEAPTTKPEAPTTKPEAPTTKPEAPTTKPEAPTTKPEAPT 678
 Db 661 PEAPPTTKAAAPNTKPEAPTTKPEAPTTKPEAPTTKPEAPTTKPEAPTTKPEAPT 720
 QY 679 APTTPKAPKELAPTTTKPEPTSTSDKAPTTKGTAPTTKPEAPTTKPEAPTTKPE 738
 Db 721 APTTPKAPKELAPTTTKPEPTSTSDKAPTTKGTAPTTKPEAPTTKPEAPTTKPE 780
 QY 739 TAPTTLKPEAPTTKPKAPKELAPTTTKGPTSTSDKAPTTKGTAPTTKPEAPTTKPE 798
 Db 781 TAPTTLKPEAPTTKPKAPKELAPTTTKGPTSTSDKAPTTKGTAPTTKPEAPTTKPE 840
 QY 799 KPAPTPTTPEPTTSEVSTPTTKPEPTTIHKSPESTPELSAETPKALENSKPEGVPT 858
 Db 841 KPAPTPTTPEPTTSEVSTPTTKPEPTTIHKSPESTPELSAETPKALENSKPEGVPT 900
 QY 859 TKTPAATKPEMTTAKOKTTERDLRTTPETTTAAPKMTKETATTTESKITATTTQV 918
 Db 901 TKTPAATKPEMTTAKOKTTERDLRTTPETTTAAPKMTKETATTTESKITATTTQV 960
 QY 919 TSTTTQDTPPKITTLKTTILAPKVTTTKITTTTEIMNKPBEETAKPKDRATNSKATPK 978
 Db 961 TSTTTQDTPPKITTLKTTILAPKVTTTKITTTTEIMNKPBEETAKPKDRATNSKATPK 1020
 QY 979 PQKPTKAPKPTSTKPKMPKPRVRAKPTTTPRKMSTMPNLPNTSRIABAMLTTRPN 1038
 Db 1021 PQKPTKAPKPTSTKPKMPKPRVRAKPTTTPRKMSTMPNLPNTSRIABAMLTTRPN 1080
 QY 1039 QTPNSKLVEVNPKSEDAGGAGETPHMLLRPHVFMPEVTPDMOVLPRVNPQGIINPMLS 1098
 Db 1091 QTPNSKLVEVNPKSEDAGGAGETPHMLLRPHVFMPEVTPDMOVLPRVNPQGIINPMLS 1140

RESULT 8

ADK67912

ID ADK67912 standard; protein; 1311 AA.

XX AC ADK67912;

XX DT 06-May-2004 (first entry)

XX

DE Human extracellular messenger (EXMES) polypeptide.
 XX Human; extracellular messenger; EXMES: respiratory-gen.; anti-allergic;
 KW anti-inflammatory; anti-infective; antidiabetic; neuroprotective;
 KW muscular-gen.; antiarthritic; osteopathic; hepatotropic; antipsoriatic;
 KW virucide; fungicide; antiparasitic; protozoacide; antihelminthic;
 XX cytotatic; gene therapy.
 OS Homo sapiens.
 XX
 FH Key Location/Qualifiers
 FT Peptide 1..30 /label= Signal_peptide
 FT /notes "Spans residues 1 to 18, 20, 21, 24, 29 or 30
 FT according to identification method"
 XX WO2004013292-A2.
 XX 12-FEB-2004.
 XX 30-JUL-2003; 2003WO-US024084.
 XX 02-AUG-2002; 2002US-0400810P.
 PR 19-SEP-2002; 2002US-0412197P.
 PR 04-OCT-2002; 2002US-0416004P.
 PR 08-NOV-2002; 2002US-0424862P.
 XX (INCY-) INCYTE CORP.
 PA Elliott VS, Khare R, Tran UK, Swarnakar A, Marquis JP,
 PI Richardson TW, Emerling BM, Lindquist EA, Chawla NK, Ramkumar J,
 PI Lee SJ;
 XX MPI; 2004-157116/15.
 DR N-PSDB; ADK67917.
 XX New extracellular messengers and nucleic acids, useful for diagnosing,
 PT treating or preventing e.g. allergies, anemia, Crohn's disease, diabetes
 PT mellitus, myasthenia gravis, osteoporosis, psoriasis, cancers or
 PT autoimmune thyroiditis.
 XX Claim 60; SEQ ID NO 5; 165pp; English.
 XX The present sequence is that of novel human extracellular messenger
 CC (EXMES) Incyte ID NO: 7513018CD1 polypeptide. The protein shows homology
 CC to human megakaryocyte stimulating factor. The invention provides EXMES
 CC polynucleotides and polypeptides, as well as expression vectors, host
 CC cells, antibodies, agonists and antagonists, and methods for diagnosing,
 CC treating or preventing disorders associated with aberrant expression of
 CC EXMES, especially autoimmune and inflammatory disorders, cell
 CC proliferative disorders and endocrine disorders, e.g. adult respiratory
 CC distress syndrome, allergy, anaemia, erythroblastosis foetalis, Crohn's
 CC disease, diabetes mellitus, myasthenia gravis, osteoarthritis,
 CC osteoporosis, pancreatitis, psoriasis, viral, bacterial, fungal,
 CC parasitic, protozoal or helminthic infections, cancers, autoimmune
 CC thyroiditis, cretinism, Plummer's disease or thyroid carcinoma.
 CC Embodiments also provide methods for using the purified EXMES and/or
 CC their encoding polynucleotides for facilitating the drug discovery
 CC process, including determining of efficacy, dosage, toxicity and
 CC pharmacology, and for investigating the pathogenesis of diseases and
 CC medical conditions.
 XX Sequence 1311 AA;
 SQ

Query Match 95.3%; Score 5607.9; DB 8; Length 1311;

Best Local Similarity 95.4%; Pred. No. 6.1e-147;

Matches 1047; Conservative 0; Mismatches 0; Indels 51; Gaps 1;

QY 1 MAWKTLPIYLLLLSVFIQQVSSQDLSSCAGRGEGYSRDATNCNDYNCQHYMECCPDF 60

Db 1 MAWKTLPIYLLLLSVFIQQVSSQDLSSCAGRGEGYSRDATNCNDYNCQHYMECCPDF 60

QY 61 KEVCTAEUSCKRCFESFERGECDDCAQCKYDKCCPDYESFCAEVHNPTSPSSKKAP 120

| | | | |
|----|------|---|------|
| Db | 61 | KRVCTAELSCGRGCFESFERGECDDAQCKYDKCCPDYESFCAE | 106 |
| Qy | 121 | PPSGASQIKSTTKSPKPPNKKTKKVI ESEETEKVKDNKQRTKKKPTPKPPVDEA | 180 |
| Db | 107 | -----VKDNKQRTKKKPTPKPPVDEA | 129 |
| Qy | 181 | GSGLDNGDFKVTTPDTSTQHNKYSTSKITAKPINRPSLPNSDTSKTSITVNET | 240 |
| Db | 130 | GSGLDNGDFKVTTPDTSTQHNKYSTSKITAKPINRPSLPNSDTSKTSITVNET | 189 |
| Qy | 241 | TVEETETTTNKQTSDDGKEKTSKAKTOSIEKTSKADLAPTSKVLAKPTPKAETTKGP | 300 |
| Db | 190 | TVEETETTTNKQTSDDGKEKTSKAKTOSIEKTSKADLAPTSKVLAKPTPKAETTKGP | 249 |
| Qy | 301 | ALTTTKEPTPTPKPPASTTKPEPTTTIKSAPTTKPEPAPTTTKSAPTTKPEPAPTTTK | 360 |
| Db | 250 | ALTTTKEPTPTPKPPASTTKPEPTTTIKSAPTTKPEPAPTTTKSAPTTKPEPAPTTTK | 309 |
| Qy | 361 | EPAPTTKPEPAPTTTKSAPTTKPEPAPTTTKKAPTTKPEPAPTTTKPEPTTK | 420 |
| Db | 310 | EPAPTTKPEPAPTTTKSAPTTKPEPAPTTTKKAPTTKPEPAPTTTKPEPTTK | 369 |
| Qy | 421 | TKPEPAPTTKPEPAPTTKAPADPKAPTTKPEPAPTTKPEPAPTTTKPEPAPTTTKPEP | 480 |
| Db | 370 | TKPEPAPTTKPEPAPTTKAPADPKAPTTKPEPAPTTTKPEPAPTTTKPEPAPTTTKPEP | 429 |
| Qy | 481 | APTITKSAPTTTKSAPTTTKSAPTTTKPEPAPTTTKPEPAPTTTKKAPPTTKP | 540 |
| Db | 430 | APTITKSAPTTTKSAPTTTKSAPTTTKPEPAPTTTKPEPAPTTTKKAPPTTKP | 489 |
| Qy | 541 | EPAPTTKPEPAPTTTKKAPTTKPEPAPTTTKETAPTTPEKLTPTTPEKLTPTTPEKAP | 600 |
| Db | 490 | EPAPTTKPEPAPTTTKKAPTTKPEPAPTTTKETAPTTPEKLTPTTPEKLTPTTPEKAP | 549 |
| Qy | 601 | TTPEELAPTTPEPTPTTPEEPAPTTPKAAAPNTKPEPAPTTKPEPAPTTKPEPAPTTK | 660 |
| Db | 550 | TTPEELAPTTPEPTPTTPEEPAPTTPKAAAPNTKPEPAPTTKPEPAPTTKPEPAPTTK | 609 |
| Qy | 661 | ETAPTTKGTAPTTKPEPAPTTKPEPAPTTTKETSTSTSDKAPTTKGTAPTTK | 720 |
| Db | 610 | ETAPTTKGTAPTTKPEPAPTTKPEPAPTTTKETSTSTSDKAPTTKGTAPTTK | 669 |
| Qy | 721 | KEPAPTTKPEPAPTTKGTAPTTKPEPAPTTKPKAPKELAPTTTKGPTSTTSKAPATT | 780 |
| Db | 670 | KEPAPTTKPEPAPTTKGTAPTTKPEPAPTTKPKAPKELAPTTTKGPTSTTSKAPATT | 729 |
| Qy | 781 | PKETAPTTKPEPAPTTKXKAPTTTETPTTSEVSTPTTKEPTTIHKSPDESTPELSA | 840 |
| Db | 730 | PKETAPTTKPEPAPTTKXKAPTTTETPTTSEVSTPTTKEPTTIHKSPDESTPELSA | 789 |
| Qy | 841 | EPTPKALENSKPEPGVPTTKIPATKPEMTTTAKDKTTERDLRTTPETTTAAAPKMTETA | 900 |
| Db | 790 | EPTPKALENSKPEPGVPTTKIPATKPEMTTTAKDKTTERDLRTTPETTTAAAPKMTETA | 849 |
| Qy | 901 | TTTEKTTESKITATTQVSTTTQDTTPPKITLKTITTLAPKVTITTKITITTEIMNKE | 960 |
| Db | 850 | TTTEKTTESKITATTQVSTTTQDTTPPKITLKTITTLAPKVTITTKITITTEIMNKE | 909 |
| Qy | 961 | ETAKPKDRATNSKATTPKOKPTKAPKKTPTSTKPKTMPKRVKPTTTPPKMTSTMPKL | 1020 |
| Db | 910 | ETAKPKDRATNSKATTPKOKPTKAPKKTPTSTKPKTMPKRVKPTTTPPKMTSTMPKL | 969 |
| Qy | 1021 | NPTTSRIAEAMLQTTTRNQTPNSKLVFNPKSDEAGGAGETPHMLLRPHVFPEVTPDM | 1080 |
| Db | 970 | NPTTSRIAEAMLQTTTRNQTPNSKLVFNPKSDEAGGAGETPHMLLRPHVFPEVTPDM | 1029 |
| Qy | 1081 | DYLPFRVNOGIIINPMLS 1098 | |
| Db | 1030 | DYLPFRVNOGIIINPMLS 1047 | |

RESULT 9

| | |
|----------|---|
| ADK65819 | |
| ID | ADK65819 standard; protein; 1320 AA. |
| XX | |
| AC | ADK65819; |
| XX | |
| DT | 06-MAY-2004 (first entry) |
| XX | |
| DE | Angiogenesis-differentially expressed protein ANH0316. |
| XX | |
| KW | cytostatic; cardiant; vasotropic; antiarteriosclerotic; |
| KW | angiogenesis inhibitor; angiogenesis stimulator; angiogenic index; |
| KW | gene expression; cancer; coronary artery disease; myocardial ischemia; |
| KW | coronary arteriosclerosis; forensic medicine. |
| XX | |
| OS | Homo sapiens. |
| XX | |
| PN | W02003066831-A2. |
| XX | |
| PD | 14-AUG-2003. |
| XX | |
| XX | 07-FEB-2003; 2003WO-US003848. |
| PF | |
| XX | 07-FEB-2002; 2002US-00067482. |
| PR | 10-JUN-2002; 2002US-00164595. |
| PR | 16-AUG-2002; 2002US-0403649P. |
| PR | 03-JAN-2003; 2003US-0437746P. |
| XX | |
| PA | (ORIG-) ORIGENE TECHNOLOGIES INC. |
| XX | |
| PI | Sun Z, Li X, Kovacs KF, Fan W, Jay G; |
| XX | |
| DR | WPI; 2003-731502/69. |
| DR | N-PSDB; ADK65818. |
| XX | |
| PT | Determining the angiogenic index of a tissue or cell sample using |
| PT | expression levels of differentially expressed genes, useful for |
| PT | diagnosing or treating cancer, coronary artery disease, myocardial |
| PT | ischemia and/or arteriosclerosis. |
| XX | |
| PS | Claim 23; SEQ ID NO 58; 296pp; English. |
| XX | |
| CC | The invention relates to a method of determining the angiogenic index of |
| CC | a tissue or cell sample comprising assessing, in a sample, the expression |
| CC | levels of one or more differentially-expressed gene from any of 34 DNA |
| CC | sequences, given in the specification, where the levels are indicative of |
| CC | the angiogenic index. The methods and compositions of the present |
| CC | invention are useful for diagnosing, preventing and/or treating cancer, |
| CC | coronary artery disease, myocardial ischemia or coronary |
| CC | arteriosclerosis. They can also be used in research, drug discovery and |
| CC | forensic medicine involving angiogenesis. This sequence corresponds to |
| CC | one of the differentially expressed proteins of the invention. |
| XX | |
| SQ | Sequence 1320 AA; |
| | Query Match 95.1%; Score 5598.8; DB 7; Length 1320; |
| | Best Local Similarity 96.0%; Pred. No. 1.1e-146; |
| | Matches 1054; Conservative 0; Mismatches 2; Indels 42; Gaps 2; |
| Qy | 1 MAWKTLPIYLLILLSVFVIQQVSSODLSSACRGCGYSRDATCNCNCOHYMECCPDF 60 |
| Db | 1 MAWKTLPIYLLILLSVFVIQQVSSQ----- |
| Qy | 61 KRVCATBELSCGRGCFESFERGECDDAQCKYDKCCPDYESFCAEVHNPTSPSSKKAP 120 |
| Db | 26 -----ELUSKGRGCFESFERGECDDAQCKYDKCCPDYESFCAEVHNPTSPSSKKAP 79 |
| Qy | 121 PPSGASQIKSTTKSPKPPNKKTKKVI ESEETEKVKDNKQRTKKKPTPKPPVDEA 180 |
| Db | 80 PPSGASQIKSTTKSPKPPNKKTKKVI ESEITE-VKDNKQRTKKKPTPKPPVDEA 138 |
| Qy | 181 GSGLDNGDFKVTTPDTSTQHNKYSTSKITAKPINRPSLPNSDTSKTSITVNET 240 |
| Db | 139 GSGLDNGDFKVTTPDTSTQHNKYSTSKITAKPINRPSLPNSDTSKTSITVNET 198 |

QY 241 TVETKETTTNNKQTSIDGKEKTTSAKETQSIEXTSAKDLAPTSKVLAKPTPKAETTTKGP 300
 Db 199 TVETKETTTNNKQTSIDGKEKTTSAKETQSIEXTSAKDLAPTSKVLAKPTPKAETTTKGP 258
 QY 301 ALTPKPEPTTTKPEPASITPKPEPTTTIKSAPTTKPEPAPTTKSAPTTKPEAPTTTK 360
 Db 259 ALTPKPEPTTTKPEPASITPKPEPTTTIKSAPTTKPEPAPTTKSAPTTKPEAPTTTK 318
 QY 361 EPAPTTTKPEPAPTTTKPEPAPTTKSAPTTKPEPAPTTKPEPAPTTKPEPAPTTTK 420
 Db 319 EPAPTTTKPEPAPTTTKPEPAPTTKSAPTTKPEPAPTTKPEPAPTTKPEPAPTTTK 378
 QY 421 TPKEPAPTTKPEPAPTTKPEPAPTTKSAPTTKPEPAPTTKPEPAPTTKPEPAPTTTK 480
 Db 379 TPKEPAPTTKPEPAPTTKPEPAPTTKSAPTTKPEPAPTTKPEPAPTTKPEPAPTTTK 438
 QY 481 APPTTKSAPTTKPEPAPTTKSAPTTKPEPAPTTKPEPAPTTKPEPAPTTKPEPAPTTTK 540
 Db 439 APPTTKSAPTTKPEPAPTTKSAPTTKPEPAPTTKPEPAPTTKPEPAPTTKPEPAPTTTK 498
 QY 541 EPAPTTTKPEPAPTTTKPEPAPTTKSAPTTKPEPAPTTKPEPAPTTKPEPAPTTTK 600
 Db 499 EPAPTTTKPEPAPTTTKPEPAPTTKSAPTTKPEPAPTTKPEPAPTTKPEPAPTTTK 558
 QY 601 TPPEELAPTTPEPTTTPEEAPTTKPEPAPTTKPEPAPTTKPEPAPTTKPEPAPTTTK 660
 Db 559 TPPEELAPTTPEPTTTPEEAPTTKPEPAPTTKPEPAPTTKPEPAPTTKPEPAPTTTK 618
 QY 661 ETAPTTTKGTAATTLKPEPAPTTKPEPAPTTKPEPAPTTKPEPAPTTKPEPAPTTTK 720
 Db 619 ETAPTTTKGTAATTLKPEPAPTTKPEPAPTTKPEPAPTTKPEPAPTTKPEPAPTTTK 678
 QY 721 KEAPTTTKPEPAPTTKPEPAPTTKPEPAPTTKPEPAPTTKPEPAPTTKPEPAPTTTK 780
 Db 679 KEAPTTTKPEPAPTTKPEPAPTTKPEPAPTTKPEPAPTTKPEPAPTTKPEPAPTTTK 738
 QY 781 PKETAPTTKPEPAPTTKPEPAPTTKPEPAPTTKPEPAPTTKPEPAPTTKPEPAPTTTK 840
 Db 739 PKETAPTTKPEPAPTTKPEPAPTTKPEPAPTTKPEPAPTTKPEPAPTTKPEPAPTTTK 798
 QY 841 EPTPKALENSPKPEPAPTTKPEPAPTTKPEPAPTTKPEPAPTTKPEPAPTTKPEPAPTTTK 900
 Db 799 EPTPKALENSPKPEPAPTTKPEPAPTTKPEPAPTTKPEPAPTTKPEPAPTTKPEPAPTTTK 858
 QY 901 TTTEKTTESKITATTTQVTSITQDTPPKITLTKTTTLPKVTTKKTTTITIMNKP 960
 Db 859 TTTEKTTESKITATTTQVTSITQDTPPKITLTKTTTLPKVTTKKTTTITIMNKP 918
 QY 961 ETAPKPKDRATNSKATTPKPKPTKAPKAPTSTKPKPTMPVRKPKTTPPKKMTSTMP 1020
 Db 919 ETAPKPKDRATNSKATTPKPKPTKAPKAPTSTKPKPTMPVRKPKTTPPKKMTSTMP 978
 QY 1021 NPTSEIAEMLQTTTRPNQNSKLVNPKSEBAGAGBETPHMLLRPHVFMPEVTPDM 1080
 Db 979 NPTSEIAEMLQTTTRPNQNSKLVNPKSEBAGAGBETPHMLLRPHVFMPEVTPDM 1038
 QY 1081 DYLPRVNPQGIINPMLS 1098
 Db 1039 DYLPRVNPQGIINPMLS 1056

RESULT 10

ADK67911

ID ADK67911 standard; protein; 1270 AA.

XX AC ADK67911;

XX DT 06-MAY-2004 (first entry)

XX DE Human extracellular messenger (EXMES); polypeptide.

XX KW Human; extracellular messenger; EXMES; respiratory-gen.; antiallergic;

KW

KW

KW

XX

XX

OS

XX

FH

FT

FT

FT

XX

XX

PN

PD

XX

PF

XX

PR

PR

PR

PR

XX

FA

XX

PI

PI

PI

XX

XX

DR

DR

XX

PT

PT

PT

XX

PS

XX

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

XX

SQ

Query Match

Best Local Similarity 90.8%; Score 5343.8; DB 8; Length 1270;

Matches 1006; Conservative 0; Mismatches 0; Indels 92; Gaps 2;

QY 1 MAWKTLPTVILLLLSVFVIQQVSSQDLSSCAGRCGEGYSRDRATNCNCOHNECCPDF 60
 |||||
 Db 1 MAWKTLPTVILLLLSVFVIQQVSSQ----- 25

QY 61 KRVCVTAELSKGRCFESPERGECDCDAQCKYDKCCPDYEFCAEVHNPTSPSSKKAP 120
 |||||
 Db 26 -----ELSKGRCFESPERGECDCDAQCKYDKCCPDYEFCAE----- 65

antiasthmatic; antiinflammatory; antidiabetic; neuroprotective;
 muscular-gen.; antiarthritic; osteopathic; hepatotropic; antipsoriatic;
 virucide; fungicide; antiparasitic; protozoacide; antihelminthic;
 cytostatic; gene therapy.

Homo sapiens.

Key Location/Qualifiers

Peptide 1..29

/label= Signal_peptide

/notes= "Spans residues 1 to 18, 20, 21, 24, 27 or 29 according to identification method"

WO2004013292-A2.

12-FEB-2004.

30-JUL-2003; 2003WO-US024084.

02-AUG-2002; 2002US-0400810P.

19-SEP-2002; 2002US-0412197P.

04-OCT-2002; 2002US-0416004P.

08-NOV-2002; 2002US-0424862P.

(INCY-) INCYTE CORP.

Elliott VS, Khare R, Tran UK, Swarnakar A, Marquis JP;

Richardson TW, Emerling BM, Lindquist EA, Chawla NK, Ramkumar J;

Lee SY;

WPI; 2004-157116/15.

N-PSDB; ADK67916.

New extracellular messengers and nucleic acids, useful for diagnosing,
 treating or preventing e.g. allergies, anemia, Crohn's disease, diabetes
 mellitus, myasthenia gravis, osteoporosis, psoriasis, cancers or
 autoimmune thyroiditis.

Claim 59; SEQ ID NO 4; 165pp; English.

The present sequence is that of novel human extracellular messenger
 (EXMES) Incyte ID NO: 7513017CD1 polypeptide. The protein shows homology
 to human megakaryocyte stimulating factor. The invention provides EXMES
 polynucleotides and polypeptides, as well as expression vectors, host
 cells, antibodies, agonists and antagonists, and methods for diagnosing,
 treating or preventing disorders associated with aberrant expression of
 EXMES, especially autoimmune and inflammatory disorders, cell
 proliferative disorders and endocrine disorders, e.g. adult respiratory
 distress syndrome, allergy, anaemia, erythroblastosis foetalis, Crohn's
 disease, diabetes mellitus, myasthenia gravis, osteoarthritis, Crohn's
 disease, osteoporosis, pancreatitis, psoriasis, viral, bacterial, fungal,
 parasitic, protozoal or helminthic infections, cancers, autoimmune
 thyroiditis, cretinism, Plummer's disease or thyroid carcinoma.
 Embodiments also provide methods for using the purified EXMES and/or
 their encoding polynucleotides for facilitating the drug discovery
 process, including determining of efficacy, dosage, toxicity and
 pharmacology, and for investigating the pathogenesis of diseases and
 medical conditions.

Sequence 1270 AA;

[illegible]

RESULT 11
ABU53252
ID ABU5
XX

CC lubricating moiety and has a polypeptide sequence comprising 1-76 repeats
CC of a motif having at least 50% identity to the sequence KEPAPTT
CC (AAB29774). The invention also relates to a nucleic acid encoding a human
CC MSF-derived tribonectin; a biocompatible composition comprising a human
CC tribonectin for inhibiting tissue adhesion formation; and a method of
CC diagnosing osteoarthritis or a predisposition to osteoarthritis by
CC measuring the amount of MSF or its fragment in a biological sample of a
CC mammal, wherein an increased amount of MSF compared to a control
CC indicates the presence of or predisposition to developing osteoarthritis.
CC The tribonectin and DNA encoding it are useful in the treatment of
CC osteoarthritis, where they may be used for lubricating mammalian joints,
CC such as articulating joints of humans, dogs or horses. The tribonectin,
CC when formulated as a membrane, foam, gel or fibre, is useful for
CC inhibiting adhesion between two surfaces such as the injured tissues of a
CC mammal, where the injury is caused by a surgical insertion or trauma, or
CC an artificial device e.g., an orthopaedic implant. In particular, one of
CC the surfaces is pericardial tissue. DNA encoding a tribonectin may be
CC used in gene therapy. The present sequence represents a substantial
CC portion of a human MSF-derived tribonectin
XX
XX Sequence 902 AA;
SQ
Query Match 48.9%; Score 2880.1; DB 4; Length 902;
Best Local Similarity 71.4%; Pred. No. 9.9e-72;
Matches 675; Conservative 27; Mismatches 74; Indels 169; Gaps 66;
QY 158 VKDNKNRTKKPTKPPVVDAGSGLDNGDFKVTTPDSTTQHKNKSVTSPIKTTAKPIN 217
DB 1 VKDNKNRTKKPTKPPVVDAGSGLDNGDFKVTTPDSTTQHKNKSVTSPIKTTAKPIN 60
QY 218 PRPSLPNSDTSKETSLSLVNKTETVETKTTTNNKQTSIDGKEKTSIAKETOSIEKTSK 277
DB 61 PRPSLPNSDTSKETSLSLVNKTETVETKTTTNNKQTSIDGKEKTSIAKETOSIEKTSK 120
QY 278 DLAPTSKVLAKPTPKAETTTKGPAL-----TTPKEPASTTPKEPTTTIKSAPTTPK 172
DB 121 DLAPTSKVLAKPTPKAETTTKGPAL-----TTPKEPASTTPKEPTTTIKSAPTTPK 172
QY 338 EPAPTTTKSAPTTPKEPAPTTKEPAPTTKEPAPTTTKSAPTTPKSAPTTPK 397
DB 173 EPAPTTTKSAPTTPKEPAP--TTKEPAPTT--KEPAP--TTKEPAPTTKEPAPTT--KEPAPTT 228
QY 398 PKPAPPTTPKEPAPTTPKAPTTPKEPAPTTKEPAPTTKEPAPTTAPKKEPAPTTKEPA 457
DB 229 --KEPAPTT--KEPAPTT--KEPAPTT--KEPAPTT--KEPAPTT--KEPAPTT--KEPA 281
QY 458 PTTTKPAPTTTKPAPTTKEPAPTTKSAPTTKSAPTTKSAPTTKSPPTTKKE 517
DB 282 PTT--KEPAP--TTKEPAPTT--KEPAPTTKEPAP--TTKEPAPTTKEPAPTT--KEPAP--TTKE 335
QY 518 PAPTTPKEPAPTTPKPAPTTKEPAPTTKEPAPTTTKKAPATAPKEPAPTTPKAPT 577
DB 336 PAPTTP--KEPAPTT--KEPAPTT--KEPAPTT--KEPAPTT--KEPAPTT--KEPAPTT 388
QY 578 TPXKLPTTPEKLAPTTPKEPAPTTPEELAPTTPEPTTTPPEAPPTPKAAAPTPKE 637
DB 389 T-KEPAPTTKEP--APTTPKE--PAPTTP--KEPAPTT--KEPAPTT--KEPAPTT--KE 440
QY 638 PAPTTPKEPAPTTKEPAPTTKETATPTPKGAPTTTLKEPAPTTPKKAP--KELAPTT 695
DB 441 PAPTTP--KEPAPTT--KEPAPTT--KEPAPTT--KEPAPTT--KEPAPTT--KEPAPTTKEPAP--T 493
QY 696 TKPPTSTSDKAPATTPKGAPTTPKEPAPTTTPKEPAPTTTKGATPTTLKEPAPTTKXP 755
DB 494 TKEPAPTTKE--PAPTTP--KEPAPTT--KEPAPTT--KEPAPTT--KEPAPTT--KEP 546
QY 756 AP--KELAPTTTKGPTSTSDKAPATTPKETATPTTPKEPAPTTPKPAPTTPTPTPTTS 813
DB 547 APTTPKEPAP--TTKEPAPTTKE--PAPTTP--KEPAPTT--KEPAPTT--KEPAPTTKE--PAPTTPK 600
QY 814 EVSTPTTTTKBPTTTTHKSPDZSTPELSAEPFPKALENSPKFPGVPTTKTAPATKPEMTT 873
DB 601 EPA--PTTKPAPTTKEPAPTT-----KEPAP-----TTKEP--APTTPKEPAPTT----- 640

QY 874 KQYTERDLRTTPTTAAAPKMTKETATTEKTESKITATTTQVTTSTTQDTPFKITT 933
DB 641 -----TKEPAPTT-----KEPAPTTKEPAPTTKEPAP----- 667
QY 934 LKTTTLAPKVTTKTKTITTEIMNKDEETAKPKDRATNSKATTPKPKQKTPAKPKPTSTK 993
DB 668 -----TTKEPAPTTKEP--APT--TK 684
QY 994 KPTMPRVKPKTPTTPRKMTSTMPBELNPTSRIZAEAMLOTTPNQTNSKLVENPKSE 1053
DB 685 EP-----APTPTPRKVTSTMPBELNPTSRIZAEAML--TTTPNQTNSKLVENPKSE 735
QY 1054 DAGGAEGETHMLLRPHVFMPEVTPDMDYLPRVFNQGIINPMLS 1098
DB 736 DAGGAEGETHMLLRPHVFMPEVTPDMDYLPRVFNQGIINPMLS 780
RESULT 14
ABUS3254
ID ABUS3254 standard; protein; 513 AA.
AC ABUS3254;
XX
XX 14-APR-2003 (first entry)
XX
XX Human testes-derived DKFZphes3_4019 homologue #3.
XX Human; gene therapy; vaccine; disease treatment; detection.
XX Homo sapiens.
XX WO200112659-A2.
XX
XX 22-FEB-2001.
XX 18-AUG-2000; 2000WO-IB001496.
XX
XX 18-AUG-1999; 99US-0149499P.
XX 28-SEP-1999; 99US-0156503P.
XX (GEHU-) GERMAN HUMAN GENOME PROJECT.
XX
XX Wiemann S;
XX
XX WPI; 2001-327840/34.
XX
XX Nucleic acids having the sequences of clones isolated from libraries of
XX different human tissues, useful in recombinant DNA methodologies.
XX
XX Example III; Page 893; 1095pp; English.
XX
XX This invention describes novel polynucleotides and polypeptides isolated
XX from human cDNA libraries which can be used for gene therapy or in
XX vaccines. The polynucleotides of the invention and antibodies encoded by
XX them may be used in the prevention, diagnosis and treatment of diseases
XX associated with inappropriate polypeptide expression. The products of the
XX invention may also be used to identify modulators of expression and
XX activity and to down regulate expression and activity. The antibodies of
XX the invention may also be used as diagnostic agents for detecting the
XX presence of polypeptides in samples. This sequence represents a homologue
XX of a polypeptide described in the disclosure of the invention
XX
SQ Sequence 513 AA;
Query Match 46.8%; Score 2757; DB 4; Length 513;
Best Local Similarity 100.0%; Pred. No. 1.2e-68;
Matches 513; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 165 RTKKKPTKPPVVDAGSGLDNGDFKVTTPDSTTQHKNKSVTSPIKTTAKPINRPSLPP 224
DB 1 RTKKKPTKPPVVDAGSGLDNGDFKVTTPDSTTQHKNKSVTSPIKTTAKPINRPSLPP 60

QY 225 NSDTSKETSIVNKTETVETKETTNTKQSTGDKETTSKETSIAKDLAPTSK 284
 DB 61 NSDTSKETSIVNKTETVETKETTNTKQSTGDKETTSKETSIAKDLAPTSK 120
 QY 285 VLAKPTPKAETTTKGPALITPKPTTTPKEPASTTPKEPTTTIKSAPTTKEPAPTTT 344
 DB 121 VLAKPTPKAETTTKGPALITPKPTTTPKEPASTTPKEPTTTIKSAPTTKEPAPTTT 180
 QY 345 KSAPTTPKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTT 404
 DB 181 KSAPTTPKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTT 240
 QY 405 TPKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTT 464
 DB 241 TPKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTT 300
 QY 465 APTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTT 524
 DB 301 APTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTT 360
 QY 525 EPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTT 584
 DB 361 EPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTT 420
 QY 585 TTPEKLAPTTTPKEPAPTTTPEELAPTTPEEPPTTPEEPPTTPEEPPTTPEEPPTTPEEPPTT 644
 DB 421 TTPEKLAPTTTPKEPAPTTTPEELAPTTPEEPPTTPEEPPTTPEEPPTTPEEPPTTPEEPPTT 480
 QY 645 EPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTT 677
 DB 481 EPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTT 513

RESULT 15

AAR80041
 ID AAR80041 standard; protein; 452 AA.
 XX
 AC AAR80041;
 XX
 DT 25-MAR-2003 (revised)
 DT 10-APR-1996 (first entry)
 XX
 DE Human megakaryocytopoietin protein.
 XX
 KW Human; megakaryocytopoietin; wheat germ agglutinin; heparin;
 KW megakaryocyte; aplastic anaemia; thrombocytopenia; hematocytopenia;
 KW multipotential stem cell.
 XX
 OS Homo sapiens.
 XX
 FH Key Location/Qualifiers
 FT Misc-difference 393..396
 FT /note= "unspecified amino acids"
 FT Misc-difference 444..446
 FT /note= "unspecified amino acids"
 FT
 FN W09523861-A1.
 XX
 PD 08-SEP-1995.
 XX
 PF 06-MAR-1995; 95WO-CN000015.
 XX
 PR 04-MAR-1994; 94CN-00112066.
 XX
 PA (SHAN-) SHANGHAI BEITE BIOTECHNOLOGY CO LTD.
 XX
 PI Gu X, Han Z, Shen Q;
 XX
 DR WPI; 1995-320576/41.
 DR N-FSD; AAT04546.
 XX
 PT New haematopoietic cell growth factor - used for treating
 PT thrombocytopenia and hematocytopenia.

XX

Example; Page 23; 36pp; Chinese.

XX This sequence represents the human megakaryocytopoietin (MPO) protein.
 CC This sequence was purified using a carrier which can couple wheat germ
 CC agglutinin and heparin to separate MPO. Fragments of this sequence (see
 CC AAR80039 and AAR80040) were used to produce the amplification primers
 CC shown in AAT04544 and AAT04545. The fragments amplified by these primers
 CC can then be used as probes to screen human cDNA libraries for MPO cDNA.
 CC The MPO cDNA can then be inserted into a plasmid which is used to
 CC transform cells to produce MPO. The MPO sequence is capable of promoting
 CC colony formation of megakaryocytes, enlarging the size of megakaryocytes
 CC and stimulating the proliferation of multipotential stem cells. The
 CC factor may be used for treating thrombocytopenia and hematocytopenia. The
 CC purification method can be used to isolate MPO from human urine or serum
 CC of patients with aplastic anaemia, and from animal blood or urine by
 CC radiation exposing the animals to induce aplastic anaemia. (Updated on 25
 CC -MAR-2003 to correct PA field.)
 XX

SQ Sequence 452 AA;

Query Match 37.8%; Score 2222.8; DB 2; Length 452;

Best Local Similarity 92.1%; Pred. No. 6.1e-54;

Matches 419; Conservative 3; Mismatches 21; Indels 12; Gaps 2;

QY 1 MAWKTLPIYLLLLSVFVIQVSSQDLSSCAGRCGEGYSRDATCNCYNCQHYMECCPDF 60

DB 1 MAWKTLPIYLLLLSVFVIQVSSQDLSSCAGRCGEGYSRDATCNCYNCQHYMECCPDF 60

QY 61 KRVTAEISCKGRCPESFERGECDDAQQKYDKCCPDYESFCABVHNPTSPSSKKAP 120

DB 61 KRVTAEISCKGRCPESFERGECDDAQQKYDKCCPDYESFCABVHNPTSPSSKKAP 120

QY 121 PPSGASQTIKSTTKRSPKPPNKKTKKVIESEBEITEKVDKNKNTKKKPTPKPPVDEA 180

DB 121 PPSGASQTIKSTTKRSPKPPNKKTKKVIESEBEITE-KVDKNKNTKKKPTPKPPVDEA 179

QY 181 GSGLDNGDFKVTTPDTSTTQHNKVTSPKITTAKPINRPSPSPNSDTSKETSIVNKT 240

DB 180 GSGLDNGDFKVTTPDTSTTQHNKVTSPKITTAKPINRPSPSPNSDTSKETSIVNKT 239

QY 241 TVETKETTNTKQSTGDKETTSKETSIAKDLAPTSKVLAKPTPKAETTTKGP 300

DB 240 TVETKETTNTKQSTGDKETTSKETSIAKDLAPTSKVLAKPTPKAETTTKGP 299

QY 301 ALTTTKEPTTTPKEPASTTPKEPTTTIKSAPTTTPKEPAPTTTTPKEPAPTTT 360

DB 300 ALTTTKEPTTTPKEPASTTPKEPTTTIKSAPTTTPKEPAPTTTTPKEPAPTTT 359

QY 361 EPAPTTTKEPAPTTTTPKEPAPTTTTPKEPAPTTTTPKEPAPTTTTPKEPAPTTT 420

DB 360 EPAPTTTKEPAPTTTTPKEPAPTTTTPKEPAPTTTTPKEPAPTTTTPKEPAPTTT 408

QY 421 TPKEPAPTTTTPKEPAPTTTTPKEPAPTTTTPKEPAPTTTTPKEPAPTTTTPKEPAPTTT 455

DB 409 LPRSLHPTTKEPAPTTTTPKEPAPTTTTPKEPAPTTTTPKEPAPTTTTPKEPAPTTT 443

Search completed: October 13, 2004, 11:36:52
 Job time : 94.5238 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: October 13, 2004, 11:23:49 ; Search time 19.0865 Seconds
(without alignments)
3815.116 Million cell updates/sec

Title: SEQ1-C
Perfect score: 5897
Sequence: 1 NAWKTLPIYLLLSVFIQ.....DMDYLPVNGIINPMLS 1098

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.1

Searched: 478139 seqs, 66318000 residues

Total number of hits satisfying chosen parameters: 478139

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued Patents AA:*
1: /cgn2_6/prodata/1/1aa/5A-COMB.pep:*
2: /cgn2_6/prodata/1/1aa/5B-COMB.pep:*
3: /cgn2_6/prodata/1/1aa/6A-COMB.pep:*
4: /cgn2_6/prodata/1/1aa/6B-COMB.pep:*
5: /cgn2_6/prodata/1/1aa/PCTUS-COMB.pep:*
6: /cgn2_6/prodata/1/1aa/backfiles1.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | ID | Description |
|------------|--------|-------------|--------|----|--------------------|
| 1 | 5872.8 | 99.8 | 1140 | 4 | US-07-757-022B-104 |
| 2 | 5872.8 | 99.8 | 1404 | 4 | US-07-757-022B-2 |
| 3 | 5872.8 | 99.8 | 1404 | 4 | US-07-757-022B-62 |
| 4 | 5872.8 | 99.8 | 1404 | 4 | US-09-298-970A-1 |
| 5 | 5871.9 | 99.7 | 1361 | 4 | US-07-757-022B-40 |
| 6 | 5863.8 | 99.6 | 1404 | 4 | US-10-164-595-78 |
| 7 | 5847.2 | 95.9 | 1364 | 4 | US-07-757-022B-48 |
| 8 | 5808.7 | 95.3 | 1363 | 4 | US-07-757-022B-52 |
| 9 | 5607.9 | 95.3 | 1311 | 4 | US-07-757-022B-42 |
| 10 | 5607.8 | 95.3 | 1320 | 4 | US-07-757-022B-46 |
| 11 | 5607.8 | 95.3 | 1320 | 4 | US-07-757-022B-60 |
| 12 | 5598.8 | 95.1 | 1320 | 4 | US-10-164-595-58 |
| 13 | 5466.9 | 92.9 | 1022 | 4 | US-07-757-022B-84 |
| 14 | 5399.2 | 91.7 | 1314 | 4 | US-07-757-022B-50 |
| 15 | 5383.1 | 91.4 | 1049 | 4 | US-07-757-022B-58 |
| 16 | 5383.1 | 91.4 | 1313 | 4 | US-07-757-022B-142 |
| 17 | 5343.8 | 90.8 | 1038 | 4 | US-07-757-022B-74 |
| 18 | 5343.8 | 90.8 | 1270 | 4 | US-07-757-022B-44 |
| 19 | 5011 | 85.1 | 941 | 4 | US-07-757-022B-14 |
| 20 | 1396.9 | 23.7 | 5179 | 4 | US-09-538-093-1258 |
| 21 | 1132.9 | 19.2 | 207 | 4 | US-07-757-022B-116 |
| 22 | 1132.9 | 19.2 | 207 | 4 | US-07-757-022B-136 |
| 23 | 1127.3 | 19.1 | 8991 | 4 | US-08-714-741-32 |
| 24 | 975.8 | 16.6 | 220 | 4 | US-07-757-022B-96 |
| 25 | 926.6 | 15.7 | 463 | 4 | US-07-757-022B-54 |
| 26 | 913.8 | 15.5 | 209 | 4 | US-07-757-022B-94 |
| 27 | 908.8 | 15.4 | 208 | 4 | US-07-757-022B-132 |

28 905.4 15.4 296 4 US-07-757-022B-70
29 899.7 15.3 231 4 US-07-757-022B-30
30 893 15.2 204 4 US-07-757-022B-92
31 886.8 15.1 192 4 US-07-757-022B-90
32 876.8 14.9 172 4 US-07-757-022B-88
33 872.5 14.8 188 4 US-07-757-022B-32
34 871 14.8 156 4 US-07-757-022B-106
35 871 14.8 156 4 US-07-757-022B-118
36 868.9 14.8 157 4 US-07-757-022B-102
37 868.9 14.8 157 4 US-07-757-022B-114
38 847.3 14.4 3256 4 US-09-919-172-98
39 847.3 14.4 3256 4 US-09-976-594-22
40 847.3 14.4 3256 4 US-09-919-039-21
41 845.9 13.9 3118 3 US-08-579-181-1
42 807.2 13.7 2972 3 US-09-579-181-2
43 800 13.6 141 4 US-07-757-022B-80
44 778 13.2 2142 4 US-09-538-092-1142
45 767.9 13.0 1837 3 US-08-928-361B-5

ALIGNMENTS

RESULT 1
US-07-757-022B-104
; Sequence 104, Application US/07757022B
; Patent No. 6433142
; GENERAL INFORMATION:
; APPLICANT: Gesner, Thomas G.
; APPLICANT: Clark, Stephen C.
; APPLICANT: Turner, Katherine
; APPLICANT: Hewick, Rodney M.
; TITLE OF INVENTION: Megakaryocyte Stimulating Factors
; NUMBER OF SEQUENCES: 143
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genetics Institute, Inc.
; STREET: 87 Cambridgepark Drive
; CITY: Cambridge
; STATE: Massachusetts
; COUNTRY: U.S.A.
; ZIP: 02140
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07757,022B
; FILING DATE: 19910910
; CLASSIFICATION: 530
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/643,502
; FILING DATE: 18-JAN-1991
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/546,114
; FILING DATE: 29-JUN-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/457,196
; FILING DATE: 29-DEC-1989
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/390,901
; FILING DATE: 08-AUG-1989
; ATTORNEY/AGENT INFORMATION:
; NAME: Cseir, Luann
; REGISTRATION NUMBER: 31,822
; REFERENCE/DOCKET NUMBER: GI 5190
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617)876-1170
; TELEFAX: (617)876-5851
; INFORMATION FOR SEQ ID NO: 104:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1140 amino acids
; TYPE: AMINO ACID

Sequence 70, Appl
Sequence 30, Appl
Sequence 92, Appl
Sequence 90, Appl
Sequence 88, Appl
Sequence 32, Appl
Sequence 106, App
Sequence 118, App
Sequence 102, App
Sequence 114, App
Sequence 98, Appl
Sequence 22, Appl
Sequence 21, Appl
Sequence 1, Appl
Sequence 2, Appl
Sequence 80, Appl
Sequence 1142, Ap
Sequence 5, Appl


```

QY 61 KVVCTAELSKGRCFESFERGECDDAQQCKYDKCCPDYESFCAEVHNPTSPSSKKAP 120
DB 61 KVVCTAELSKGRCFESFERGECDDAQQCKYDKCCPDYESFCAEVHNPTSPSSKKAP 120
QY 121 PPASGASQIKSTTKRSPKPNKKTKKVIIESEITE----- 156
DB 121 PPASGASQIKSTTKRSPKPNKKTKKVIIESEITE----- 180
QY 157 -----KVNDKKNRTKKKTKKVIIESEITE----- 198
DB 157 -----KVNDKKNRTKKKTKKVIIESEITE----- 240
QY 181 KIKSKNSAANRELQKLLKVKDNKKNRTKKKTKKVIIESEITE----- 258
DB 181 KIKSKNSAANRELQKLLKVKDNKKNRTKKKTKKVIIESEITE----- 300
QY 199 TOHNVSTSPKITTAKPINRPSLDPNSDTSKBSLTVNKESTTVETKETTINKQTSIDG 258
DB 199 TOHNVSTSPKITTAKPINRPSLDPNSDTSKBSLTVNKESTTVETKETTINKQTSIDG 300
QY 241 TOHNVSTSPKITTAKPINRPSLDPNSDTSKBSLTVNKESTTVETKETTINKQTSIDG 300
DB 241 TOHNVSTSPKITTAKPINRPSLDPNSDTSKBSLTVNKESTTVETKETTINKQTSIDG 318
QY 259 KEKTSASAKTOSIETSAKDLAPTSKVLAKPTPKAETTTKGPALTTKPEPTTTPKBPAS 318
DB 259 KEKTSASAKTOSIETSAKDLAPTSKVLAKPTPKAETTTKGPALTTKPEPTTTPKBPAS 360
QY 301 KEKTSASAKTOSIETSAKDLAPTSKVLAKPTPKAETTTKGPALTTKPEPTTTPKBPAS 360
DB 301 KEKTSASAKTOSIETSAKDLAPTSKVLAKPTPKAETTTKGPALTTKPEPTTTPKBPAS 378
QY 319 TPKKPTTTTKSAPTTKBPAPTTTKGAPTTKBPAPTTTKGAPTTKBPAPTTTKGAPTT 378
DB 319 TPKKPTTTTKSAPTTKBPAPTTTKGAPTTKBPAPTTTKGAPTTKBPAPTTTKGAPTT 420
QY 361 TPKKPTTTTKSAPTTKBPAPTTTKGAPTTKBPAPTTTKGAPTTKBPAPTTTKGAPTT 420
DB 361 TPKKPTTTTKSAPTTKBPAPTTTKGAPTTKBPAPTTTKGAPTTKBPAPTTTKGAPTT 438
QY 379 APTTTKSAPTTKBPAPTTKBPAPTTKBPAPTTKBPAPTTKBPAPTTKBPAPTTKBP 438
DB 379 APTTTKSAPTTKBPAPTTKBPAPTTKBPAPTTKBPAPTTKBPAPTTKBPAPTTKBP 480
QY 421 APTTTKSAPTTKBPAPTTKBPAPTTKBPAPTTKBPAPTTKBPAPTTKBPAPTTKBP 480
DB 421 APTTTKSAPTTKBPAPTTKBPAPTTKBPAPTTKBPAPTTKBPAPTTKBPAPTTKBP 498
QY 439 EPAPTAPKKPAPTTKBPAPTTKBPAPTTKBPAPTTKBPAPTTKBPAPTTKBPAPT 498
DB 439 EPAPTAPKKPAPTTKBPAPTTKBPAPTTKBPAPTTKBPAPTTKBPAPTTKBPAPT 540
QY 481 EPAPTAPKKPAPTTKBPAPTTKBPAPTTKBPAPTTKBPAPTTKBPAPTTKBPAPT 540
DB 481 EPAPTAPKKPAPTTKBPAPTTKBPAPTTKBPAPTTKBPAPTTKBPAPTTKBPAPT 558
QY 499 TTKSAPTTKBPAPTTKBPAPTTKBPAPTTKBPAPTTKBPAPTTKBPAPTTKBPAPT 558
DB 499 TTKSAPTTKBPAPTTKBPAPTTKBPAPTTKBPAPTTKBPAPTTKBPAPTTKBPAPT 600
QY 541 TTKSAPTTKBPAPTTKBPAPTTKBPAPTTKBPAPTTKBPAPTTKBPAPTTKBPAPT 600
DB 541 TTKSAPTTKBPAPTTKBPAPTTKBPAPTTKBPAPTTKBPAPTTKBPAPTTKBPAPT 618
QY 559 APTAKEPAPTTKBPAPTTKBPAPTTKBPAPTTKBPAPTTKBPAPTTKBPAPTPTT 618
DB 559 APTAKEPAPTTKBPAPTTKBPAPTTKBPAPTTKBPAPTTKBPAPTTKBPAPTPTT 660
QY 601 APTAKEPAPTTKBPAPTTKBPAPTTKBPAPTTKBPAPTTKBPAPTTKBPAPTPTT 660
DB 601 APTAKEPAPTTKBPAPTTKBPAPTTKBPAPTTKBPAPTTKBPAPTTKBPAPTPTT 678
QY 619 PEPAPTTKBPAPTTKBPAPTTKBPAPTTKBPAPTTKBPAPTTKBPAPTTKBPAPT 678
DB 619 PEPAPTTKBPAPTTKBPAPTTKBPAPTTKBPAPTTKBPAPTTKBPAPTTKBPAPT 720
QY 661 PEPAPTTKBPAPTTKBPAPTTKBPAPTTKBPAPTTKBPAPTTKBPAPTTKBPAPT 720
DB 661 PEPAPTTKBPAPTTKBPAPTTKBPAPTTKBPAPTTKBPAPTTKBPAPTTKBPAPT 738
QY 679 APTTKKAPKELAPTTKBPAPTTKBPAPTTKBPAPTTKBPAPTTKBPAPTTKBPAPT 738
DB 679 APTTKKAPKELAPTTKBPAPTTKBPAPTTKBPAPTTKBPAPTTKBPAPTTKBPAPT 780
QY 721 APTTKKAPKELAPTTKBPAPTTKBPAPTTKBPAPTTKBPAPTTKBPAPTTKBPAPT 780
DB 721 APTTKKAPKELAPTTKBPAPTTKBPAPTTKBPAPTTKBPAPTTKBPAPTTKBPAPT 798
QY 739 TAPTTLKBPAPTTKBPAPTTKBPAPTTKBPAPTTKBPAPTTKBPAPTTKBPAPTTPK 798
DB 739 TAPTTLKBPAPTTKBPAPTTKBPAPTTKBPAPTTKBPAPTTKBPAPTTKBPAPTTPK 840
QY 781 TAPTTLKBPAPTTKBPAPTTKBPAPTTKBPAPTTKBPAPTTKBPAPTTKBPAPTTPK 840
DB 781 TAPTTLKBPAPTTKBPAPTTKBPAPTTKBPAPTTKBPAPTTKBPAPTTKBPAPTTPK 858
QY 799 KPAPTTPPTTSEVSTPTTKBPAPTTKBPAPTTKBPAPTTKBPAPTTKBPAPTTKBPAPT 858
DB 799 KPAPTTPPTTSEVSTPTTKBPAPTTKBPAPTTKBPAPTTKBPAPTTKBPAPTTKBPAPT 900
QY 841 KPAPTTPPTTSEVSTPTTKBPAPTTKBPAPTTKBPAPTTKBPAPTTKBPAPTTKBPAPT 900
DB 841 KPAPTTPPTTSEVSTPTTKBPAPTTKBPAPTTKBPAPTTKBPAPTTKBPAPTTKBPAPT 918
QY 859 TKTPAATKEMITTAADKTTERTDLRTTPTTTTAAKQWTKETATTTKTESITATTTQV 918
DB 859 TKTPAATKEMITTAADKTTERTDLRTTPTTTTAAKQWTKETATTTKTESITATTTQV 960
QY 901 TKTPAATKEMITTAADKTTERTDLRTTPTTTTAAKQWTKETATTTKTESITATTTQV 960
DB 901 TKTPAATKEMITTAADKTTERTDLRTTPTTTTAAKQWTKETATTTKTESITATTTQV 978
QY 919 TSTTTQDTPPKITTLKTTTLAPKVTTKKTTTTHIMNKPEETAKPKDRATNSKATTPK 978
DB 919 TSTTTQDTPPKITTLKTTTLAPKVTTKKTTTTHIMNKPEETAKPKDRATNSKATTPK 1020
QY 961 TSTTTQDTPPKITTLKTTTLAPKVTTKKTTTTHIMNKPEETAKPKDRATNSKATTPK 1020
DB 961 TSTTTQDTPPKITTLKTTTLAPKVTTKKTTTTHIMNKPEETAKPKDRATNSKATTPK 1038
QY 979 PQKPTKAPKPTSTKPKPTMVRVKPTTTPRKTSTWPNLNTSRIAEAMLQTTTRPN 1038
DB 979 PQKPTKAPKPTSTKPKPTMVRVKPTTTPRKTSTWPNLNTSRIAEAMLQTTTRPN 1080
QY 1021 PQKPTKAPKPTSTKPKPTMVRVKPTTTPRKTSTWPNLNTSRIAEAMLQTTTRPN 1080
DB 1021 PQKPTKAPKPTSTKPKPTMVRVKPTTTPRKTSTWPNLNTSRIAEAMLQTTTRPN 1098
QY 1039 QTPNSKLVENPKSEDAGGAETPHMLLRPHVFPVETPDMDYLRVFNQGIINPMLS 1098
DB 1039 QTPNSKLVENPKSEDAGGAETPHMLLRPHVFPVETPDMDYLRVFNQGIINPMLS 1140
QY 1081 QTPNSKLVENPKSEDAGGAETPHMLLRPHVFPVETPDMDYLRVFNQGIINPMLS 1140
DB 1081 QTPNSKLVENPKSEDAGGAETPHMLLRPHVFPVETPDMDYLRVFNQGIINPMLS

```

```

RESULT 3
US-07-757-022B-62
; Sequence 62, Application US/07757022B
; Patent No. 6433142
; GENERAL INFORMATION:
; APPLICANT: Gesner, Thomas G.
; APPLICANT: Clark, Stephen C.
; APPLICANT: Turner, Katherine
; APPLICANT: Hewick, Rodney M.
; TITLE OF INVENTION: Megakaryocyte Stimulating Factors
; NUMBER OF SEQUENCES: 143
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genetics Institute, Inc.
; STREET: 87 CambridgePark Drive
; CITY: Cambridge
; STATE: Massachusetts
; COUNTRY: U.S.A.
; ZIP: 02140
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/757,022B
; FILING DATE: 19910910
; CLASSIFICATION: 530
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/643,502
; FILING DATE: 18-JAN-1991
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/546,114
; FILING DATE: 29-JUN-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/457,196
; FILING DATE: 29-DEC-1989
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/390,901
; FILING DATE: 08-AUG-1989
; ATTORNEY/AGENT INFORMATION:
; NAME: Cserk, Luann
; REGISTRATION NUMBER: 31,822
; REFERENCE/DOCKET NUMBER: GI 5190
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617)876-1170
; TELEFAX: (617)876-5851
; INFORMATION FOR SEQ ID NO: 62:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1404 amino acids
; TYPE: AMINO ACID
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-07-757-022B-62

```

```

Query Match 99.88; Score 5872.8; DB 4; Length 1404;
Best Local Similarity 96.33; Pred. No. 5:5e-171;
Matches 1098; Conservative 0; Mismatches 0; Indels 42; Gaps 1;

QY 1 MAWKTLPIYLLLSVFIQVSSQDLSSCAGRCGEGYSDATCNCNDYQCHYMECCPDF 60
DB 1 MAWKTLPIYLLLSVFIQVSSQDLSSCAGRCGEGYSDATCNCNDYQCHYMECCPDF 60
QY 61 KRVTAELSCKGRCFESFERGECDDAQQCKYDKCCPDYESFCAEVHNPTSPSSKKAP 120
DB 61 KRVTAELSCKGRCFESFERGECDDAQQCKYDKCCPDYESFCAEVHNPTSPSSKKAP 120
QY 121 PPASGASQIKSTTKRSPKPNKKTKKVIIESEITE----- 156
DB 121 PPASGASQIKSTTKRSPKPNKKTKKVIIESEITE----- 180
QY 157 -----KVNDKKNRTKKKTKKVIIESEITE----- 198
DB 157 -----KVNDKKNRTKKKTKKVIIESEITE-----

```

| | | | |
|----|------|--|------|
| Db | 181 | KIKSSKNSAANRELQKKUKLVKDNNKQNRKKKPTPKPPVVVDAGSGLDNGDFKVITPDNST | 241 |
| Qy | 199 | TOHNKVSZPKITTAQINPRPSPSPNSDTSKETSILTVMKETTIVTKETTTTTNKQTSIDG | 258 |
| Db | 241 | TOHNKVSZPKITTAQINPRPSPSPNSDTSKETSILTVMKETTIVTKETTTTTNKQTSIDG | 300 |
| Qy | 259 | KEKTTSAKETOSIEKTSAKOLAPTSKVLAKPTPKAEITTTKGPALTTPKEPTPTPKCEPAS | 318 |
| Db | 301 | KEKTTSAKETOSIEKTSAKOLAPTSKVLAKPTPKAEITTTKGPALTTPKEPTPTPKCEPAS | 360 |
| Qy | 319 | TTPEPTPTTTKSAPTTPKECAPTTTKSAPTTTPKECAPTTTKCAPTTTPKCPAPTTTKCP | 378 |
| Db | 361 | TTPEPTPTTTKSAPTTPKECAPTTTKSAPTTTPKECAPTTTKCAPTTTPKCPAPTTTKCP | 420 |
| Qy | 379 | APTITKSAPTTPKBPAPTTPKKAPTTPKCBAPTTTPKBPOTTTTPKECAPTTTKCAPTTTKP | 438 |
| Db | 421 | APTITKSAPTTPKBPAPTTPKKAPTTTPKBPAPTTTPKBPOTTTTPKECAPTTTKCAPTTTKP | 480 |
| Qy | 439 | EPAPTAPKKPAPTTTPKECAPTTTPKECAPTTTKKESPTTPKECAPTTTKSAPTTTKECAP | 498 |
| Db | 481 | EPAPTAPKKPAPTTTPKECAPTTTPKECAPTTTKKESPTTPKECAPTTTKSAPTTTKECAP | 540 |
| Qy | 499 | TTKSAPTTPKBPSTTTTPKECAPTTTPKECAPTTTPKKAPTTTPKECAPTTTPKBPAPTTTKP | 558 |
| Db | 541 | TTKSAPTTPKBPSTTTTPKECAPTTTPKBPAPTTTPKKAPTTTPKECAPTTTPKBPAPTTTKP | 600 |
| Qy | 559 | APTAPKEBPAPTTPKETAPTTTPKKLTPTTPPEKLAPTTPPEKAPTTTPPELAPTTTPPEPT | 618 |
| Db | 601 | APTAPKEBPAPTTPKETAPTTTPKKLTPTTPPEKLAPTTPPEKAPTTTPPELAPTTTPPEPT | 660 |
| Qy | 619 | PEEPAPTTPKAAAPNTPKCAPTTTPKCAPTTTPKCAPTTTPKETAPTTPKGTAPTTTLKEP | 678 |
| Db | 661 | PEEPAPTTPKAAAPNTPKCAPTTTPKCAPTTTPKCAPTTTPKETAPTTPKGTAPTTTLKEP | 720 |
| Qy | 679 | APTTPKBPAPKELAPTTTKBPTSTTSOKPAPTTTPKGTAPTTTPKBPAPTTTPKBPAPTTPKG | 738 |
| Db | 721 | APTTPKBPAPKELAPTTTKBPTSTTSOKPAPTTTPKGTAPTTTPKBPAPTTTPKBPAPTTPKG | 780 |
| Qy | 739 | TAPTTLKEAPTTTPKBPAPKELAPTTTKGPTSTTSDKPAPTTTPKETAPTTPKBPAPTTPK | 798 |
| Db | 781 | TAPTTLKEAPTTTPKBPAPKELAPTTTKGPTSTTSDKPAPTTTPKETAPTTPKBPAPTTPK | 840 |
| Qy | 799 | KPAPTTPEPTPTTSEVSTPTTTTKEPTTIHKSPOSSTPPELSAETTPKALENSPKEGVP | 858 |
| Db | 841 | KPAPTTPEPTPTTSEVSTPTTTTKEPTTIHKSPOSSTPPELSAETTPKALENSPKEGVP | 900 |
| Qy | 859 | TKTPAAATKPMTTTTAKDKTTTERDLRTTPEITTTAAAPKMTKETATTTETKTTSSKITATTQV | 918 |
| Db | 901 | TKTPAAATKPMTTTTAKDKTTTERDLRTTPEITTTAAAPKMTKETATTTETKTTSSKITATTQV | 960 |
| Qy | 919 | TSITTTQDTPFKITTLTKITTLAPKVTTTTKTIITTEIMNKPEETAKEPKDRATNSKATTPK | 978 |
| Db | 961 | TSITTTQDTPFKITTLTKITTLAPKVTTTTKTIITTEIMNKPEETAKEPKDRATNSKATTPK | 1020 |
| Qy | 979 | PQKPTKAPKAPTSTKKPKTPVRVKPKTTPTPKQWSTMPPELNPSTSIABAMLOTTTRPN | 1038 |
| Db | 1021 | PQKPTKAPKAPTSTKKPKTPVRVKPKTTPTPKQWSTMPPELNPSTSIABAMLOTTTRPN | 1080 |
| Qy | 1039 | QTPNSKLVFNPKSDEAGGAEGTTPHMLLRPHVPMPEVTTPMDYLPVPVNOGIIINPMLS | 1098 |
| Db | 1081 | QTPNSKLVFNPKSDEAGGAEGTTPHMLLRPHVPMPEVTTPMDYLPVPVNOGIIINPMLS | 1140 |

RESULT 4

```

US-09-298-970A-1
; Sequence 1, Application US/09298970A
; Patent No. 6743774
; GENERAL INFORMATION:
; APPLICANT: JAY, Gregory D.
; TITLE OF INVENTION: TRIBONECTINS
; FILE REFERENCE: 21486-026
; CURRENT APPLICATION NUMBER: US/09/298,970A
; CURRENT FILING DATE: 2001-06-19

```

Db 901 TKTPAATPEMTTAKDKTTERDLRTTPEITTAAPKMTKETATTEKTTESKITATTQV 960
 Qy 919 TSTTTQDTPTEFKITLTKTTLAPKVTTTKKTIITTEIWNKPEETAKPKDRATNSKATTPK 978
 Db 961 TSTTTQDTPTEFKITLTKTTLAPKVTTTKKTIITTEIWNKPEETAKPKDRATNSKATTPK 1020
 Qy 979 POKPTKAPKPTSTKPKTMPRVKPKTPTTPRQMTSTWPELNPTSRIAEAMLTTPRN 1038
 Db 1021 POKPTKAPKPTSTKPKTMPRVKPKTPTTPRQMTSTWPELNPTSRIAEAMLTTPRN 1080
 Qy 1039 QTPNSKLVENPKSDAGAGETPHMLLRPHVFPVETTPDMXYLPRVNOGIINPMLS 1098
 Db 1081 QTPNSKLVENPKSDAGAGETPHMLLRPHVFPVETTPDMXYLPRVNOGIINPMLS 1140

RESULT 5
 US-07-757-022B-40
 ; Sequence 40, Application US/07757022B
 ; Patent No. 6433142
 ; GENERAL INFORMATION:
 ; APPLICANT: Gesner, Thomas G.
 ; APPLICANT: Clark, Stephen C.
 ; APPLICANT: Turner, Katherine
 ; APPLICANT: Hewick, Rodney M.
 ; TITLE OF INVENTION: Megakaryocyte Stimulating Factors
 ; NUMBER OF SEQUENCES: 143
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Genetics Institute, Inc.
 ; STREET: 87 Cambridgepark Drive
 ; CITY: Cambridge
 ; STATE: Massachusetts
 ; COUNTRY: U.S.A.
 ; ZIP: 02140
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: PatentIn Release #1.0, Version #1.25
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/07/757,022B
 ; FILING DATE: 19910910
 ; CLASSIFICATION: 530
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: US 07/643,502
 ; FILING DATE: 18-JAN-1991
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: US 07/546,114
 ; FILING DATE: 29-JUN-1990
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: US 07/457,196
 ; FILING DATE: 29-DEC-1989
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: US 07/390,901
 ; FILING DATE: 08-AUG-1989
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Cserr, Luann
 ; REGISTRATION NUMBER: 31,822
 ; REFERENCE/DOCKET NUMBER: GI 5190
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: (617) 876-1170
 ; TELEFAX: (617) 876-5851
 ; INFORMATION FOR SEQ ID NO: 40:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 1361 amino acids
 ; TYPE: AMINO ACID
 ; TOPOLOGY: linear
 ; MOLECULE TYPE: protein
 ; US-07-757-022B-40

Query Match 99.7%; Score 5871.9; DB 4; Length 1361;
 Best Local Similarity 99.9%; Pred. No. 5.6e-171;
 Matches 1097; Conservative 0; Mismatches 0; Indels 1; Gaps 1;

Qy 1 MAWKTLPIYLLLLSVFVIQVSSQDLSSCAGRCGEGYGRDATCNCDYNCQHWMECCPDF 60
 Db 1 MAWKTLPIYLLLLSVFVIQVSSQDLSSCAGRCGEGYGRDATCNCDYNCQHWMECCPDF 60
 Qy 61 KRVTAEALSCKGRCFSEFERGRCDDAOCKYDKCCPDYSESCAEVHNPTSPSSKAP 120
 Db 61 KRVTAEALSCKGRCFSEFERGRCDDAOCKYDKCCPDYSESCAEVHNPTSPSSKAP 120
 Qy 121 PPSGASQTIKSTTKRSPKPPNKTKKVIIESEBITEKVKDNKNKRTKKKPTPKPPVUDEA 180
 Db 121 PPSGASQTIKSTTKRSPKPPNKTKKVIIESEBITE - VKDNKNKRTKKKPTPKPPVUDEA 179
 Qy 181 GSGLDNGDFKVTTPDSTTQHNKVSPKTIITAKPINRPSLPPNSDTSKETSLSLVNKET 240
 Db 180 GSGLDNGDFKVTTPDSTTQHNKVSPKTIITAKPINRPSLPPNSDTSKETSLSLVNKET 239
 Qy 241 TVETKETTTNNKOTSIDGKEKTTSAKETSIAKTSADLAPTSKVLAKPTPKAEATTTKGP 300
 Db 240 TVETKETTTNNKOTSIDGKEKTTSAKETSIAKTSADLAPTSKVLAKPTPKAEATTTKGP 299
 Qy 301 ALATPKPTPTTPKBPASTTPKPTPTTTIKSAPTTKPEAPATTTKSAPTTKPEPATTTTK 360
 Db 300 ALATPKPTPTTPKBPASTTPKPTPTTTIKSAPTTKPEAPATTTKSAPTTKPEPATTTTK 359
 Qy 361 EPAPTTPKBPATTTKPEAPATTTKSAPTTPKPEAPATTTKPEAPATTTKPEPTPT 420
 Db 360 EPAPTTPKBPATTTKPEAPATTTKSAPTTPKPEAPATTTKPEAPATTTKPEPTPT 419
 Qy 421 TPKEPATTTKPEAPATTTKPEAPATTTKPEAPATTTKPEAPATTTKPEPTTPKEP 480
 Db 420 TPKEPATTTKPEAPATTTKPEAPATTTKPEAPATTTKPEAPATTTKPEPTTPKEP 479
 Qy 481 APITTSAPTTKPEAPATTTKSAPTTPKPSPTTKPEAPATTTKPEAPATTTKPEPATTPK 540
 Db 480 APITTSAPTTKPEAPATTTKSAPTTPKPSPTTKPEAPATTTKPEAPATTTKPEPATTPK 539
 Qy 541 EPAPTTPKBPATTTKPEAPATTTKPEAPATTTKPEAPATTTKPEAPATTTKPEPAT 600
 Db 540 EPAPTTPKBPATTTKPEAPATTTKPEAPATTTKPEAPATTTKPEAPATTTKPEPAT 599
 Qy 601 TTPEELAPTTPEPTPTTPEEPATTPKAAAPNTPKAPATTTKPEAPATTTKPEPATTPK 660
 Db 600 TTPEELAPTTPEPTPTTPEEPATTPKAAAPNTPKAPATTTKPEAPATTTKPEPATTPK 659
 Qy 661 ETAPTTKPGTAPTTKPEAPATTTKPEAPATTTKPEPTSTTSDKAPATTTKPGTAPTT 720
 Db 660 ETAPTTKPGTAPTTKPEAPATTTKPEAPATTTKPEPTSTTSDKAPATTTKPGTAPTT 719
 Qy 721 KEPAATTPKBPATTPKGTAPTTKPEAPATTTKPEAPATTTKPEPTSTTSDKAPATTT 780
 Db 720 KEPAATTPKBPATTPKGTAPTTKPEAPATTTKPEAPATTTKPEPTSTTSDKAPATTT 779
 Qy 781 PKETAPTTKPEAPATTTKPEAPATTTPEPTTSEVSTPTTKETTTIKSPDSTPELSA 840
 Db 780 PKETAPTTKPEAPATTTKPEAPATTTPEPTTSEVSTPTTKETTTIKSPDSTPELSA 839
 Qy 841 EPTPKALENSPKPEGVPTTKTTPAATKPEMTTTAKDTERDLRTTPTTTTAAAPKMTKETA 900
 Db 840 EPTPKALENSPKPEGVPTTKTTPAATKPEMTTTAKDTERDLRTTPTTTTAAAPKMTKETA 899
 Qy 901 TTTTEKTTESKITATTTQVSTTTTODTTPFKITLTKTTLAPKVTTTKKTIITTEIWNKPE 960
 Db 900 TTTTEKTTESKITATTTQVSTTTTODTTPFKITLTKTTLAPKVTTTKKTIITTEIWNKPE 959
 Qy 961 ETAKPKDRATNSKATTPKPKPTKAPKPTSTKPKTMPRVKPKTPTTPRQMTSTWPELN 1020
 Db 960 ETAKPKDRATNSKATTPKPKPTKAPKPTSTKPKTMPRVKPKTPTTPRQMTSTWPELN 1019
 Qy 1021 NPTSRIAEAMLTTPRNQTPNSKLVENPKSDAGAGETPHMLLRPHVFPVETTPDM 1080
 Db 1020 NPTSRIAEAMLTTPRNQTPNSKLVENPKSDAGAGETPHMLLRPHVFPVETTPDM 1079


```

; REFERENCE/DOCKET NUMBER: GI 5190
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617)876-1170
; TELEFAX: (617)876-5851
; INFORMATION FOR SEQ ID NO: 48:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1354 amino acids
; TYPE: AMINO ACID
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-07-757-022B-48

Query Match          95.9%; Score 5647.2; DB 4; Length 1354;
Best Local Similarity 96.3%; Pred. No. 3.8e-164;
Matches 1057; Conservative 7; Mismatches 26; Indels 8; Gaps 1;

QY 1 MAWKTLPIYLLLLSVFIQVSSQDLSSCAGRCGEGYSDATNCDCYNCOHYMECCPDF 60
DB 1 MAWKTLPIYLLLLSVFIQVSSQDLSSCAGRCGEGYSDATNCDCYNCOHYMECCPDF 60

QY 61 KRVCTAEISCKGRCTESFERGECDCDAQCKYDKCCPDYSEFCAEVHNPTSPSSKKAP 120
DB 61 KRVCTAEISCKGRCTESFERGECDCDAQCKYDKCCPDYSEFCAEVHNPTSPSSSSS 120

QY 121 PPSGASQITKSTTKSPKPPNKKTKVIBSEITEKVDNKNKNTKKKPTKPPVDEA 180
DB 121 SSSSSSTWIKSSKNSAARELOKLL-----KVDNKNKNTKKKPTKPPVDEA 172

QY 181 GSGLONGPQVTTPTSTQHNKVTSPKLTAKPINRPSLPNPSDTSKETSIVNKET 240
DB 173 GSGLONGPQVTTPTSTQHNKVTSPKLTAKPINRPSLPNPSDTSKETSIVNKET 232

QY 241 TVETKETTITNKQSTDGKEKTSIAKETSIAKOLAPTSKVLAKPTPKAETTTKGP 300
DB 233 TVETKETTITNKQSTDGKEKTSIAKETSIAKOLAPTSKVLAKPTPKAETTTKGP 292

QY 301 ALTTPEKPTPTPKPASTTPKEPTTIKSAPTTPKEPAPTTTKSAPTTPKEPAPTTK 360
DB 293 ALTTPEKPTPTPKPASTTPKEPTTIKSAPTTPKEPAPTTTKSAPTTPKEPAPTTK 352

QY 361 EPAPTTPEKAPTTTKAPTTTKSAPTTPKPAPTTPKKAPTTPKPAPTTPKPAPT 420
DB 353 EPAPTTPEKAPTTTKAPTTTKSAPTTPKPAPTTPKKAPTTPKPAPTTPKPAPT 412

QY 421 TPKEPAPTTKEPAPTTKEPAPTTAPKAPPTTPKEPAPTTTPKEPAPTTTKPSP 480
DB 413 TPKEPAPTTKEPAPTTKEPAPTTAPKAPPTTPKEPAPTTTPKEPAPTTTKPSP 472

QY 481 APPTTKSAPTTKEPAPTTTKSAPTTPKPAPTTTPKEPAPTTTPKEPAPTTPK 540
DB 473 APPTTKSAPTTKEPAPTTTKSAPTTPKPAPTTTPKEPAPTTTPKEPAPTTPK 532

QY 541 EPAPTTPEKAPTTTKKAPTAPEPAPTTTPKETAPTTPKLTPTTPEKLAPTTPEK 600
DB 533 EPAPTTPEKAPTTTKKAPTAPEPAPTTTPKETAPTTPKLTPTTPEKLAPTTPEK 592

QY 601 TPEELAPPTPEPTTTEEPAPTTPKAAANTPKPAPTTPKPAPTTTPKEPAPTTPK 660
DB 593 TPEELAPPTPEPTTTEEPAPTTPKAAANTPKPAPTTPKPAPTTTPKEPAPTTPK 652

QY 661 ETAPTTPKGTAPTTLKEPAPTTPKDAPKELAPTTTKPTSTTSOKPAPTTTKGAPT 720
DB 653 ETAPTTPKGTAPTTLKEPAPTTPKDAPKELAPTTTKPTSTTSOKPAPTTTKGAPT 712

QY 721 KEAPTTPEKAPTTTPKGTAPTTLKEPAPTTPKDAPKELAPTTTKGPTSTTSOK 780
DB 713 KEAPTTPEKAPTTTPKGTAPTTLKEPAPTTPKDAPKELAPTTTKGPTSTTSOK 772

QY 781 PKETAPTTPEKAPTTTPKAPTTTPETPPPTTSEVSTPTTKEPTTIHKSPPDES 840
DB 773 PKETAPTTPEKAPTTTPKAPTTTPETPPPTTSEVSTPTTKEPTTIHKSPPDES 832

QY 841 EPTPKALENSPKPGVPTTKTTPAATKPEMTTTAKDKTTERDLRTTPTTAAAPKMTKETA 900

```

```

RESULT 8
US-07-757-022B-52
; Sequence 52, Application US/07757022B
; Patent No. 6433142
; GENERAL INFORMATION:
; APPLICANT: Gesner, Thomas G.
; APPLICANT: Clark, Stephen C.
; APPLICANT: Turner, Katherine
; APPLICANT: Hewick, Rodney M.
; TITLE OF INVENTION: Megakaryocyte Stimulating Factors
; NUMBER OF SEQUENCES: 143
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genetics Institute, Inc.
; STREET: 87 Cambridgepark Drive
; CITY: Cambridge
; STATE: Massachusetts
; COUNTRY: U.S.A.
; ZIP: 02140
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/757,022B
; FILING DATE: 19910910
; CLASSIFICATION: 530
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/643,502
; FILING DATE: 18-JAN-1991
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/546,114
; FILING DATE: 29-JUN-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/457,196
; FILING DATE: 29-DEC-1989
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/390,901
; FILING DATE: 08-AUG-1989
; ATTORNEY/AGENT INFORMATION:
; NAME: Cseri, Luann
; REGISTRATION NUMBER: 31,822
; REFERENCE/DOCKET NUMBER: GI 5190
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617)876-1170
; TELEFAX: (617)876-5851
; INFORMATION FOR SEQ ID NO: 52:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1363 amino acids
; TYPE: AMINO ACID
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-07-757-022B-52

```


Db 61 KRVCTAELSCGRCFESFRRGRCDCDAQCKYDKCCPDYBSFCAE----- 106
Qy 121 PPASGASQTIKSTTKSPKPNKKTKKVIKIESEBETKVKDNKNKNTKKKTPKPPVVDVA 180
Db 107 -----VKDNKNKNTKKKTPKPPVVDVA 129
Qy 181 GSGLDNGDFKVTTPDSTTQHNKVSSTPKITTAKEPINRPSLPPNSDTSKETSLSLVNKET 240
Db 130 GSGLDNGDFKVTTPDSTTQHNKVSSTPKITTAKEPINRPSLPPNSDTSKETSLSLVNKET 189
Qy 241 TVETKETTITNNKQSTGDKKEKTSIAKETSIEKTSIAKOLAPTSKVLAKPTPKAETTTKGP 300
Db 190 TVETKETTITNNKQSTGDKKEKTSIAKETSIEKTSIAKOLAPTSKVLAKPTPKAETTTKGP 249
Qy 301 ALTTKPEPTTTPKBPASTTKEPTPTTKSAPTTKBPAPTTTKSAPTTKBPAPTTTK 360
Db 250 ALTTKPEPTTTPKBPASTTKEPTPTTKSAPTTKBPAPTTTKSAPTTKBPAPTTTK 309
Qy 361 EPAPTTKBPAPTTTKBPAPTTTKGAPTTKBPAPTTTKKAPTTKBPAPTTTKPEPTT 420
Db 310 EPAPTTKBPAPTTTKBPAPTTTKGAPTTKBPAPTTTKKAPTTKBPAPTTTKPEPTT 369
Qy 421 TPKEPAPTTKBPAPTTKBPAPTTAKKAPTTKBPAPTTTKBPAPTTTKBPAPTTTK 480
Db 370 TPKEPAPTTKBPAPTTKBPAPTTAKKAPTTKBPAPTTTKBPAPTTTKBPAPTTTK 429
Qy 481 APTTTKSAPTTTKBPAPTTKAPTTKBPAPTTTKBPAPTTTKBPAPTTTKBPAPTTK 540
Db 430 APTTTKSAPTTTKBPAPTTKAPTTKBPAPTTTKBPAPTTTKBPAPTTTKBPAPTTK 489
Qy 541 EPAPTTKBPAPTTTKKAPTTAKKAPTTKBPAPTTTKKAPTTTKKAPTTTKPEKAP 600
Db 490 EPAPTTKBPAPTTTKKAPTTAKKAPTTKBPAPTTTKKAPTTTKKAPTTTKPEKAP 549
Qy 601 TTPELAPTTPEPTTPEBPAPTTPKAAANTKBPAPTTKBPAPTTTKBPAPTTTKBPAPTTK 660
Db 550 TTPELAPTTPEPTTPEBPAPTTPKAAANTKBPAPTTKBPAPTTTKBPAPTTTKBPAPTTK 609
Qy 661 ETAPTTPKGATPTTKBPAPTTPKKAPKELAPTTTKPTSTTSKAPTTTKGAPTTT 720
Db 610 ETAPTTPKGATPTTKBPAPTTPKKAPKELAPTTTKPTSTTSKAPTTTKGAPTTT 669
Qy 721 KBPAPTTKBPAPTTKBPAPTTTLKBPAPTTPKKAPKELAPTTTKGPTSTTSKAPTT 780
Db 670 KBPAPTTKBPAPTTKBPAPTTTLKBPAPTTPKKAPKELAPTTTKGPTSTTSKAPTT 729
Qy 781 PKETAPTTKBPAPTTKBPAPTTTETPPPTTSEVSTPTTKPTTTHKSPDESPELSA 840
Db 730 PKETAPTTKBPAPTTKBPAPTTTETPPPTTSEVSTPTTKPTTTHKSPDESPELSA 789
Qy 841 EPTPKALENSPKBPVPTTKTPAATKPEMTTAKDKTTERDLRTTPTTTAAPKMTKETA 900
Db 790 EPTPKALENSPKBPVPTTKTPAATKPEMTTAKDKTTERDLRTTPTTTAAPKMTKETA 849
Qy 901 TTTKTTESKLTATTTQVSTTTQDTTPKITTTLKTTTLAPKVTTKKTTTTEIMNKP 960
Db 850 TTTKTTESKLTATTTQVSTTTQDTTPKITTTLKTTTLAPKVTTKKTTTTEIMNKP 909
Qy 961 ETAKPKDRATNSKATTPKPKTKAPKPTSTKPKPTWPRVAKPTTTPTRKMTSTMP 1020
Db 910 ETAKPKDRATNSKATTPKPKTKAPKPTSTKPKPTWPRVAKPTTTPTRKMTSTMP 969
Qy 1021 NPTSRIAEAMLOTTTRPNQTPNSKLVEVNPKSEDAGGAEGETPHMLLRPHVMPVPTDM 1080
Db 970 NPTSRIAEAMLOTTTRPNQTPNSKLVEVNPKSEDAGGAEGETPHMLLRPHVMPVPTDM 1029
Qy 1081 DYLPRVNOGIIINPMLS 1098
Db 1030 DYLPRVNOGIIINPMLS 1047

RESULT 10

US-07-757-022B-46

/ Sequence 46, Application US/07757022B
/ Patent No. 6433142
/ GENERAL INFORMATION:
/ APPLICANT: Gesner, Thomas G.
/ APPLICANT: Clark, Stephen C.
/ APPLICANT: Turner, Katherine
/ APPLICANT: Hewick, Rodney M.
/ TITLE OF INVENTION: Megakaryocyte Stimulating Factors
/ NUMBER OF SEQUENCES: 143
/ CORRESPONDENCE ADDRESS:
/ ADDRESSEE: Genetics Institute, Inc.
/ STREET: 87 CambridgePark Drive
/ CITY: Cambridge
/ STATE: Massachusetts
/ COUNTRY: U.S.A.
/ ZIP: 02140
/ COMPUTER READABLE FORM:
/ MEDIUM TYPE: Floppy disk
/ COMPUTER: IBM PC compatible
/ OPERATING SYSTEM: PC-DOS/MS-DOS
/ SOFTWARE: PatentIn Release #1.0, Version #1.25
/ CURRENT APPLICATION DATA:
/ APPLICATION NUMBER: US/07/757,022B
/ FILING DATE: 19910910
/ CLASSIFICATION: 530
/ PRIOR APPLICATION DATA:
/ APPLICATION NUMBER: US 07/643,502
/ FILING DATE: 18-JAN-1991
/ PRIOR APPLICATION DATA:
/ APPLICATION NUMBER: US 07/546,114
/ FILING DATE: 29-JUN-1990
/ PRIOR APPLICATION DATA:
/ APPLICATION NUMBER: US 07/457,196
/ FILING DATE: 29-DEC-1989
/ PRIOR APPLICATION DATA:
/ APPLICATION NUMBER: US 07/390,901
/ FILING DATE: 08-AUG-1989
/ ATTORNEY/AGENT INFORMATION:
/ NAME: Cserr, Luann
/ REGISTRATION NUMBER: 31,822
/ REFERENCE/DOCKET NUMBER: GI 5190
/ TELECOMMUNICATION INFORMATION:
/ TELEPHONE: (617)876-1170
/ TELEFAX: (617)876-5851
/ INFORMATION FOR SEQ ID NO: 46:
/ SEQUENCE CHARACTERISTICS:
/ LENGTH: 1320 amino acids
/ TYPE: AMINO ACID
/ TOPOLOGY: linear
/ MOLECULE TYPE: protein
/ US-07-757-022B-46

Query Match 95.3%; Score 5607.8; DB 4; Length 1320;
Best Local Similarity 96.2%; Pred. No. 5.8e-163;
Matches 1056; Conservative 0; Mismatches 0; Indels 42; Gaps 2;

Qy 1 MANKTLPIYLLLLSVFVIQQVSSQDLSSCAGRGEGYSRDATAICNDYCNQHYNECCPDF 60
Db 1 MANKTLPIYLLLLSVFVIQQVSSQ----- 25
Qy 61 KRVCTAELSCGRCFESFRRGRCDCDAQCKYDKCCPDYBSFCAEVHNPTSPSSKKAP 120
Db 26 -----ELSCGRCFESFRRGRCDCDAQCKYDKCCPDYBSFCAEVHNPTSPSSKKAP 79
Qy 121 PPASGASQTIKSTTKSPKPNKKTKKVIKIESEBETKVKDNKNKNTKKKTPKPPVVDVA 180
Db 80 PPASGASQTIKSTTKSPKPNKKTKKVIKIESEBETKVKDNKNKNTKKKTPKPPVVDVA 138
Qy 181 GSGLDNGDFKVTTPDSTTQHNKVSSTPKITTAKEPINRPSLPPNSDTSKETSLSLVNKET 240
Db 139 GSGLDNGDFKVTTPDSTTQHNKVSSTPKITTAKEPINRPSLPPNSDTSKETSLSLVNKET 198
Qy 241 TVETKETTITNNKQSTGDKKEKTSIAKETSIEKTSIAKOLAPTSKVLAKPTPKAETTTKGP 300

```

199 TVETKETTNNKQSTDCGKNTSASKEQTSIEKTSKADLAPTSKVLAKPTKAEITTKGP 258
301 ALTTKKEPTTTPKEPASSTPKKEPTTTIKSAPTTKEPAPTTTKSAPTTKEPAPTTTK 360
259 ALTTKKEPTTTPKEPASSTPKKEPTTTIKSAPTTKEPAPTTTKSAPTTKEPAPTTTK 318
361 EPAPTTKEPAPTTTKKEPAPTTTKSAPTTKEPAPTTTKKAPTTTPKEPAPTTTKKEPTT 420
319 EPAPTTKEPAPTTTKKEPAPTTTKSAPTTKEPAPTTTPKAPTTTPKEPAPTTTKKEPTT 378
421 TPKEPAPTTKGPAPTTTPKEPAPTTAPKAPAPTTTPKEPAPTTTPKEPAPTTTKKEPTT 480
379 TPKEPAPTTKGPAPTTTPKEPAPTTAPKAPAPTTTPKEPAPTTTPKEPAPTTTKKEPTT 438
481 APPTTKSAPTTTKKEPAPTTTKSAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTKKEPTT 540
439 APPTTKSAPTTTKKEPAPTTTKSAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTKKEPTT 498
541 EPAPTTKEPAPTTTKKAPAPTTAPKAPAPTTTPKEPAPTTTPKLTPTTTPKAPTTTPKEPAP 600
499 EPAPTTKEPAPTTTKKAPAPTTAPKAPAPTTTPKEPAPTTTPKLTPTTTPKAPTTTPKEPAP 558
601 TPPELAPTTPEEPTTTPKEPAPTTTPKAAAPNTPKAPAPTTTPKEPAPTTTPKEPAPTTTK 660
559 TPPELAPTTPEEPTTTPKEPAPTTTPKAAAPNTPKAPAPTTTPKEPAPTTTPKEPAPTTTK 618
661 ETAPTTTPKGTAPTTTLKEPAPTTTPKAPAPKELAPTTTPKEPTTSTGDKAPAPTTTPKGTATTP 720
619 ETAPTTTPKGTAPTTTLKEPAPTTTPKAPAPKELAPTTTPKEPTTSTGDKAPAPTTTPKGTATTP 678
721 KEAPAPTTKEPAPTTTPKGTATPTTLKEPAPTTTPKAPAPKELAPTTTPKGTSTTSKAPAPTT 780
679 KEAPAPTTKEPAPTTTPKGTATPTTLKEPAPTTTPKAPAPKELAPTTTPKGTSTTSKAPAPTT 738
781 PKETAPTTTPKAPAPTTTPKAPAPTTTPETPTTTPSEVSTPTTTPKEPTTIHKSDESTPBLSA 840
739 PKETAPTTTPKAPAPTTTPKAPAPTTTPETPTTTPSEVSTPTTTPKEPTTIHKSDESTPBLSA 798
841 EPTPKALENSKPEGPVPTTKTAPAKPMTTAKDKTTERDLRTTPTTAAAPKMTKETA 900
799 EPTPKALENSKPEGPVPTTKTAPAKPMTTAKDKTTERDLRTTPTTAAAPKMTKETA 858
901 TTTETKTESKITATTTQVTSITQDTPFKITTLKTTTLAPKVTITTKTITTTIMKNKE 960
859 TTTETKTESKITATTTQVTSITQDTPFKITTLKTTTLAPKVTITTKTITTTIMKNKE 918
961 ETAPKPKDRATNSKATTPKPKPTKAPKXPTSTKPKMTPRVRKPKTTTPTRKMTSTMPKL 1020
919 ETAPKPKDRATNSKATTPKPKPTKAPKXPTSTKPKMTPRVRKPKTTTPTRKMTSTMPKL 978
1021 NPTSIARAMLQTTTRPNQTPNSKLVEYNPKSEBAGGAGETPHMLLAPHVFMPEVTPDM 1080
979 NPTSIARAMLQTTTRPNQTPNSKLVEYNPKSEBAGGAGETPHMLLAPHVFMPEVTPDM 1038
1081 DYLPVPVPGQIINPMLS 1098
1039 DYLPVPVPGQIINPMLS 1056

```

RESULT 11

```

US-07-757-022B-60
; Sequence 60, Application US/07757022B
; Patent No. 6433142
; GENERAL INFORMATION:
; APPLICANT: Gesner, Thomas G.
; APPLICANT: Clark, Stephen C.
; APPLICANT: Turner, Katherine
; APPLICANT: Hewick, Rodney M.
; TITLE OF INVENTION: Megakaryocyte Stimulating Factors
; NUMBER OF SEQUENCES: 143
; CORRESPONDENCE ADDRESS:
; ADDRESSSEE: Genetics Institute, Inc.

```

```

; STREET: 87 CambridgePark Drive
; CITY: Cambridge
; STATE: Massachusetts
; COUNTRY: U.S.A.
; ZIP: 02140
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/757,022B
; FILING DATE: 19910910
; CLASSIFICATION: 530
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/643,502
; FILING DATE: 18-JAN-1991
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/546,114
; FILING DATE: 29-JUN-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/457,196
; FILING DATE: 29-DEC-1989
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/390,901
; FILING DATE: 08-AUG-1989
; ATTORNEY/AGENT INFORMATION:
; NAME: Cserr, Luann
; REGISTRATION NUMBER: 31,822
; REFERENCE/DOCKET NUMBER: GI 5190
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617)876-1170
; TELEFAX: (617)876-5851
; INFORMATION FOR SEQ ID NO: 60:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1320 amino acids
; TYPE: AMINO ACID
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-07-757-022B-60

```

```

Query Match 95.3%; Score 5607.8; DB 4; Length 1320;
Best Local Similarity 96.2%; Pred. No. 5.8e-163;
Matches 1056; Conservative 0; Mismatches 0; Indels 42; Gaps 2;

```

```

QY 1 MAWKTLPIYLLLLSVFVIQVSSQDLSCAGCGEGYSRDATCNCYDNCQHYMECCPDF 60
DB 1 MAWKTLPIYLLLLSVFVIQVSSQ----- 25
QY 61 KRVCYTAELSCKGRCPESFERGECDCDAQCKKYDKCCPDYBSFCAEVHNPSPSSKKAP 120
DB 26 -----ELSCKGRCPESFERGECDCDAQCKKYDKCCPDYBSFCAEVHNPSPSSKKAP 79
QY 121 PPSGASQTIKSTTKSPKPPNKKTKKVIKESIEIKVKDKNTKKTTPKPPVDEA 180
DB 80 PPSGASQTIKSTTKSPKPPNKKTKKVIKESIEITE-VKDNKKNTKKKTPKPPVDEA 138
QY 181 GSGLDNGDFKVTTPDTSTTQHNKSVTSKITTAKINPRPSLPNSDTSKETSIVNKET 240
DB 139 GSGLDNGDFKVTTPDTSTTQHNKSVTSKITTAKINPRPSLPNSDTSKETSIVNKET 198
QY 241 TVETKETTNNKQSTDCGKNTSASKEQTSIEKTSKADLAPTSKVLAKPTKAEITTKGP 300
DB 199 TVETKETTNNKQSTDCGKNTSASKEQTSIEKTSKADLAPTSKVLAKPTKAEITTKGP 258
QY 301 ALTTKKEPTTTPKBPASTTPKEPTTTIKSAPTTKEPAPTTTKSAPTTTPKEPAPTTTK 360
DB 259 ALTTKKEPTTTPKBPASTTPKEPTTTIKSAPTTTPKEPAPTTTKSAPTTTPKEPAPTTTK 318
QY 361 EPAPTTTPKEPAPTTTKKEPAPTTTKSAPTTTPKEPAPTTTPKAPAPTTTPKEPAPTTTK 420
DB 319 EPAPTTTPKEPAPTTTKKEPAPTTTKSAPTTTPKEPAPTTTPKAPAPTTTPKEPAPTTTK 378

```


RESULT 13
 US-07-757-022B-84
 ; Sequence 84, Application US/07757022B
 ; Patent No. 6433142
 ; GENERAL INFORMATION:
 ; APPLICANT: Gesner, Thomas G.
 ; APPLICANT: Clark, Stephen C.
 ; APPLICANT: Turner, Katherine
 ; APPLICANT: Hewick, Rodney M.
 ; TITLE OF INVENTION: Megakaryocyte Stimulating Factors
 ; NUMBER OF SEQUENCES: 143
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Genetics Institute, Inc.
 ; STREET: 87 CambridgePark Drive
 ; CITY: Cambridge
 ; STATE: Massachusetts
 ; COUNTRY: U.S.A.
 ; ZIP: 02140
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: Patent in Release #1.0, Version #1.25
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/07757,022B
 ; FILING DATE: 19910910
 ; CLASSIFICATION: 530
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: US 07/643,502
 ; FILING DATE: 18-JAN-1991
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: US 07/546,114
 ; FILING DATE: 29-JUN-1990
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: US 07/457,196
 ; FILING DATE: 29-DEC-1989
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: US 07/390,901
 ; FILING DATE: 08-AUG-1989
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Csert, Luann
 ; REGISTRATION NUMBER: 31,822
 ; REFERENCE/DOCKET NUMBER: GI 5190
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: (617) 876-1170
 ; TELEFAX: (617) 876-5851
 ; INFORMATION FOR SEQ ID NO: 84:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 1022 amino acids
 ; TYPE: AMINO ACID
 ; TOPOLOGY: linear
 ; MOLECULE TYPE: protein
 ; US-07-757-022B-84

Query Match 92.9%; Score 5466.9; DB 4; Length 1022;
 Best Local Similarity 95.1%; Pred. No. 8e-159;
 Matches 1020; Conservative 0; Mismatches 2; Indels 51; Gaps 2;
 QY 26 DLSSCAGCGEGYGRDATNCNDYNCQHYMECCPDKRVCTAELSCKGRCFESFERGRECD 85
 DB 1 DLSSCAGCGEGYGRDATNCNDYNCQHYMECCPDKRVCTAELSCKGRCFESFERGRECD 58
 QY 86 CDAOCKYDKCCPYESCAEVHNPSPSSKXAPPPSGASQITKSTTKSPXPPNKKY 145
 DB 59 CDAOCKYDKCCPYESCAEVHNPSPSSKXAPPPSGASQITKSTTKSPXPPNKKY 80
 QY 146 KKVIESEITEKVDKNKNTKKKTPKPPVVDAGSLGNDGPKVTTPTDSTTQHNKVS 205
 DB 81 -----AVKDNKKNRTKKKTPKPPVVDAGSLGNDGPKVTTPTDSTTQHNKVS 129
 QY 206 TSPKLTAKPINRPSPNSDTSKTSLVNKETTVEKTTTNNKQSTDGKETTSA 265

DB 130 TSPKLTAKPINRPSPNSDTSKTSLVNKETTVEKTTTNNKQSTDGKETTSA 189
 QY 266 KETQSIEKTSKDLAPTSKVLAKPTPKAETTTKGPALTTTPKEPTTTPKEPASTTTPKEPT 325
 DB 190 KETQSIEKTSKDLAPTSKVLAKPTPKAETTTKGPALTTTPKEPTTTPKEPASTTTPKEPT 249
 QY 326 PTTIKSAPTTPKEPAPTTTKASPTTKKEPAPTTTKKEPAPTTTKKEPAPTTTKS 385
 DB 250 PTTIKSAPTTPKEPAPTTTKASPTTKKEPAPTTTKKEPAPTTTKKEPAPTTTKS 309
 QY 386 APTTPKEPAPTTTPKXPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAP 445
 DB 310 APTTPKEPAPTTTPKXPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAP 369
 QY 446 KKPAPTTTPKEPAPTTTKSPPTTKSPPTTKSPPTTKSPPTTKSPPTTKSPPT 505
 DB 370 KKPAPTTTPKEPAPTTTKSPPTTKSPPTTKSPPTTKSPPTTKSPPTTKSPPT 429
 QY 506 TPKEPSPTTKKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKE 565
 DB 430 TPKEPSPTTKKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKE 489
 QY 566 PAPTTPKETAPTTPKXLTPTTPEKLAPTTPEKAPTTPEELAPTTPEELAPTTPEELAP 625
 DB 490 PAPTTPKETAPTTPKXLTPTTPEKLAPTTPEKAPTTPEELAPTTPEELAPTTPEELAP 549
 QY 626 TPKAAAPNTPKGPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKE 685
 DB 550 TPKAAAPNTPKGPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKE 609
 QY 686 PAPKELAPTTTKEPTSTTSKDPAPTTTPKGTAPTTPKEPAPTTTPKEPAPTTTPKEPAP 745
 DB 610 PAPKELAPTTTKEPTSTTSKDPAPTTTPKGTAPTTPKEPAPTTTPKEPAPTTTPKEPAP 669
 QY 746 EPAPTTTPKXPAPKELAPTTTKGPTSTTSKDPAPTTTPKGTAPTTPKEPAPTTTPKEPAP 805
 DB 670 EPAPTTTPKXPAPKELAPTTTKGPTSTTSKDPAPTTTPKGTAPTTPKEPAPTTTPKEPAP 729
 QY 806 ETPTPTTSEVSTPTTKEPTTIHKSPDESTPSELSAETPKALENSPKGPGVPTTKTPAAT 865
 DB 730 ETPTPTTSEVSTPTTKEPTTIHKSPDESTPSELSAETPKALENSPKGPGVPTTKTPAAT 789
 QY 866 KPEMTTAKDKTTERDLRTPPTTAAAPKATKATTTTEKTTESKITATTQVTTSTTQD 925
 DB 790 KPEMTTAKDKTTERDLRTPPTTAAAPKATKATTTTEKTTESKITATTQVTTSTTQD 849
 QY 926 TTPFKITLLKTTTLAPKVTTKKTTTTEIMNKPEETAKPKDRATNSKATTPKPKPTKA 985
 DB 850 TTPFKITLLKTTTLAPKVTTKKTTTTEIMNKPEETAKPKDRATNSKATTPKPKPTKA 909
 QY 986 PKKPTSTKKPKTMRVRKPKTTTPPKMTSTNPELNPTSRIAEAMLTQTTTRPNQTPNSKL 1045
 DB 910 PKKPTSTKKPKTMRVRKPKTTTPPKMTSTNPELNPTSRIAEAMLTQTTTRPNQTPNSKL 969
 QY 1046 VEVNPKSEDAGGAGETPHMLRPHVFMPEVTPDMDYLPRVFNQGIILINPMLS 1098
 DB 970 VEVNPKSEDAGGAGETPHMLRPHVFMPEVTPDMDYLPRVFNQGIILINPMLS 1022

RESULT 14
 US-07-757-022B-50
 ; Sequence 50, Application US/07757022B
 ; Patent No. 6433142
 ; GENERAL INFORMATION:
 ; APPLICANT: Gesner, Thomas G.
 ; APPLICANT: Clark, Stephen C.
 ; APPLICANT: Turner, Katherine
 ; APPLICANT: Hewick, Rodney M.
 ; TITLE OF INVENTION: Megakaryocyte Stimulating Factors
 ; NUMBER OF SEQUENCES: 143
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Genetics Institute, Inc.

```

; STREET: 87 Cambridgepark Drive
; CITY: Cambridge
; STATE: Massachusetts
; COUNTRY: U.S.A.
; ZIP: 02140
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US 07/757,022B
; FILING DATE: 19910910
; CLASSIFICATION: 530
; PRIORITY APPLICATION DATA:
; APPLICATION NUMBER: US 07/643,502
; FILING DATE: 18-JAN-1991
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/546,114
; FILING DATE: 29-JUN-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/457,196
; FILING DATE: 29-DEC-1989
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/390,901
; FILING DATE: 08-AUG-1989
; ATTORNEY/AGENT INFORMATION:
; NAME: Cseir, Luann
; REGISTRATION NUMBER: 31,822
; REFERENCE/DOCKET NUMBER: GI 5190
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617)876-1170
; TELEFAX: (617)876-5851
; INFORMATION FOR SEQ ID NO: 50:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1314 amino acids
; TYPE: AMINO ACID
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-07-757-022B-50

Query Match          91.7%; Score 5399.2; DB 4; Length 1314;
Best Local Similarity 92.7%; Pred. No. 1.3e-156;
Matches 1018; Conservative 7; Mismatches 25; Indels 48; Gaps 2;

QY   1  MAWKTLPIYLLLLSVFVIQVSSQDLSSCAGCGEGYSRDATCNCYNCQHYNECCPDF 60
Db   1  MAWKTLPIYLLLLSVFVIQVSSQDLSSCAGCGEGYSRDATCNCYNCQHYNECCPDF 60

QY   61  KRVTAEALSCKGRCFESFERGRECDCAOCKKYDKCCPDYESFCAEVHNPTSPPPSSKKAP 120
Db   61  KRVTAE-----HSVSENQSSSSS 80

QY   121  PPSGASQTKSTTKSPKPPNKKTKKVIIESEITEKVDNKNQTKKKPTPKPPVUDEA 180
Db   81  SSSSSSTTKWIKSSKNSAANLEQLK-----KVDNKNRTKKKPTPKPPVUDEA 132

QY   181  GSGLDNGDFKVTTPDSTTQHNKVSPTSITAKPINRPSLPNSDTSKETSLSLVNKET 240
Db   133  GSGLDNGDFKVTTPDSTTQHNKVSPTSITAKPINRPSLPNSDTSKETSLSLVNKET 192

QY   241  TVETKETTNTKQTSFDGKEKTSKETSIEKTSIAKDLAPTSKVLAKPTPKAETTKGP 300
Db   193  TVETKETTNTKQTSFDGKEKTSKETSIEKTSIAKDLAPTSKVLAKPTPKAETTKGP 252

QY   301  ALTTTPKEPTTPKPEASTTPKEPTPTTKSAPTTPKEAPTTPKSAPTTPKPEAPTTPK 360
Db   253  ALTTTPKEPTTPKPEASTTPKEPTPTTKSAPTTPKPEAPTTPKSAPTTPKPEAPTTPK 312

QY   361  EPAPTTTPKEAPTTPKPEAPTTPKSAPTTPKPEAPTTPKSAPTTPKPEAPTTPK 420
Db   313  EPAPTTTPKEAPTTPKPEAPTTPKSAPTTPKPEAPTTPKSAPTTPKPEAPTTPK 372

; STREET: 87 Cambridgepark Drive
; CITY: Cambridge
; STATE: Massachusetts
; COUNTRY: U.S.A.
; ZIP: 02140
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US 07/757,022B
; FILING DATE: 19910910
; CLASSIFICATION: 530
; PRIORITY APPLICATION DATA:
; APPLICATION NUMBER: US 07/643,502
; FILING DATE: 18-JAN-1991
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/546,114
; FILING DATE: 29-JUN-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/457,196
; FILING DATE: 29-DEC-1989
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/390,901
; FILING DATE: 08-AUG-1989
; ATTORNEY/AGENT INFORMATION:
; NAME: Cseir, Luann
; REGISTRATION NUMBER: 31,822
; REFERENCE/DOCKET NUMBER: GI 5190
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617)876-1170
; TELEFAX: (617)876-5851
; INFORMATION FOR SEQ ID NO: 50:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1314 amino acids
; TYPE: AMINO ACID
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-07-757-022B-50

RESULT 15
US-07-757-022B-58
; Sequence 58, Application US/07757022B
; Patent No. 6433142
; GENERAL INFORMATION:
; APPLICANT: Gesner, Thomas G.
; APPLICANT: Clark, Stephen C.
; APPLICANT: Turner, Katherine
; APPLICANT: Hewick, Rodney M.
; TITLE OF INVENTION: Megakaryocyte Stimulating Factors
; NUMBER OF SEQUENCES: 143
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genetics Institute, Inc.
; STREET: 87 Cambridgepark Drive
; CITY: Cambridge
; STATE: Massachusetts
; COUNTRY: U.S.A.
; ZIP: 02140
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:

```


GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: October 13, 2004, 11:23:49 ; Search time 111.709 Seconds

(without alignments)
3171.696 Million cell updates/sec

Title: SEQ1-C

Perfect score: 5887

Sequence: 1 MAWKTLPIYLLLSVFIQ.....DMDYLPVFNQGIINPMLS 1098

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.1

Searched: 1356558.seqs, 322682953 residues

Total number of hits satisfying chosen parameters: 1356558

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications AA:*

1: /cgn2_6/ptodata/1/pubpaa/US07_PUBCOMB.pep:*

2: /cgn2_6/ptodata/1/pubpaa/PCT_NEW_PUB.pep:*

3: /cgn2_6/ptodata/1/pubpaa/US05_NEW_PUB.pep:*

4: /cgn2_6/ptodata/1/pubpaa/US06_PUBCOMB.pep:*

5: /cgn2_6/ptodata/1/pubpaa/US07_NEW_PUB.pep:*

6: /cgn2_6/ptodata/1/pubpaa/PCTUS_PUBCOMB.pep:*

7: /cgn2_6/ptodata/1/pubpaa/US08_NEW_PUB.pep:*

8: /cgn2_6/ptodata/1/pubpaa/US08_PUBCOMB.pep:*

9: /cgn2_6/ptodata/1/pubpaa/US09A_PUBCOMB.pep:*

10: /cgn2_6/ptodata/1/pubpaa/US09B_PUBCOMB.pep:*

11: /cgn2_6/ptodata/1/pubpaa/US09C_PUBCOMB.pep:*

12: /cgn2_6/ptodata/1/pubpaa/US09_NEW_PUB.pep:*

13: /cgn2_6/ptodata/1/pubpaa/US10A_PUBCOMB.pep:*

14: /cgn2_6/ptodata/1/pubpaa/US10C_PUBCOMB.pep:*

15: /cgn2_6/ptodata/1/pubpaa/US10D_PUBCOMB.pep:*

16: /cgn2_6/ptodata/1/pubpaa/US10_NEW_PUB.pep:*

17: /cgn2_6/ptodata/1/pubpaa/US60_NEW_PUB.pep:*

18: /cgn2_6/ptodata/1/pubpaa/US60_PUBCOMB.pep:*

19: /cgn2_6/ptodata/1/pubpaa/US60_PUBCOMB.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | ID | Description |
|------------|--------|-------------|--------|----|-------------------|
| 1 | 5872.8 | 99.8 | 1140 | 13 | US-10-124-557-104 |
| 2 | 5872.8 | 99.8 | 1404 | 9 | US-09-802-207-30 |
| 3 | 5872.8 | 99.8 | 1404 | 11 | US-09-897-188-1 |
| 4 | 5872.8 | 99.8 | 1404 | 13 | US-10-124-557-2 |
| 5 | 5872.8 | 99.8 | 1404 | 13 | US-10-124-557-62 |
| 6 | 5871.9 | 99.7 | 1361 | 13 | US-10-124-557-40 |
| 7 | 5647.2 | 95.9 | 1354 | 13 | US-10-124-557-48 |
| 8 | 5608.7 | 95.3 | 1363 | 13 | US-10-124-557-52 |
| 9 | 5607.9 | 95.3 | 1311 | 13 | US-10-124-557-42 |
| 10 | 5607.8 | 95.3 | 1320 | 13 | US-10-124-557-46 |
| 11 | 5607.8 | 95.3 | 1320 | 13 | US-10-124-557-60 |
| 12 | 5466.9 | 92.9 | 1022 | 13 | US-10-124-557-84 |
| 13 | 5399.2 | 91.7 | 1314 | 13 | US-10-124-557-50 |
| 14 | 5383.1 | 91.4 | 1049 | 13 | US-10-124-557-58 |

15 5383.1 91.4 1313 13 US-10-124-557-142

16 5383.8 90.8 1038 13 US-10-124-557-74

17 5343.8 90.6 1270 13 US-10-124-557-44

18 5011 85.1 941 13 US-10-124-557-14

19 2508.6 42.6 792 9 US-09-802-207-27

20 1566.8 26.6 292 16 US-10-468-910-4

21 1409.9 23.9 538 14 US-10-038-694-3

22 1396.9 23.7 5179 9 US-09-922-217-1068

23 1396.9 23.7 5179 9 US-09-933-263-1068

24 1396.9 23.7 5179 13 US-10-025-380-1068

25 1396.9 23.7 5179 16 US-10-734-564-121

26 1132.9 19.2 207 13 US-10-124-557-116

27 1132.9 19.2 207 13 US-10-124-557-116

28 1025.8 17.4 185 16 US-10-468-910-2

29 1009.9 17.2 188 14 US-10-038-694-2

30 975.8 16.6 220 13 US-10-124-557-96

31 968.2 16.4 1460 14 US-10-295-027-428

32 949.8 16.1 1325 9 US-09-864-761-35612

33 926.6 15.7 463 13 US-10-124-557-54

34 923.9 15.7 1367 9 US-09-801-368-108

35 913.8 15.5 209 13 US-10-124-557-94

36 910.5 15.5 3507 14 US-10-369-493-5784

37 908.8 15.4 208 13 US-10-124-557-132

38 905.4 15.4 296 13 US-10-124-557-70

39 899.7 15.3 231 13 US-10-124-557-30

40 893 15.2 204 13 US-10-124-557-92

41 886.8 15.1 192 13 US-10-124-557-90

42 876.8 14.9 172 13 US-10-124-557-88

43 872.9 14.8 6642 14 US-10-369-493-5013

44 872.5 14.8 188 13 US-10-124-557-32

45 871 14.8 156 13 US-10-124-557-106

ALIGNMENTS

RESULT 1

US-10-124-557-104
; Sequence 104, Application US/10124557
; Publication No. US20020137894A1
; GENERAL INFORMATION:
; APPLICANT: Turner, Katherine
; Clark, Stephen C.
; Jacobs, Kenneth
; Hewick, Rodney M.
; Gesner, Thomas G.
; TITLE OF INVENTION: Megakaryocyte Stimulating Factors
; NUMBER OF SEQUENCES: 143
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genetics Institute, Inc.
; STREET: 87 Cambridgepark Drive
; CITY: Cambridge
; STATE: Massachusetts
; COUNTRY: U.S.A.
; ZIP: 02140
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/10/124.557
; FILING DATE: 16-Apr-2002
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/643,502
; FILING DATE: 18-JAN-1991
; APPLICATION NUMBER: US 07/546,114
; FILING DATE: 29-JUN-1990
; APPLICATION NUMBER: US 07/457,196
; FILING DATE: 29-DEC-1989
; APPLICATION NUMBER: US 07/390,901
; FILING DATE: 08-AUG-1989

ATTORNEY/AGENT INFORMATION:
 NAME: Cseri, Luann
 REGISTRATION NUMBER: 31,822
 REFERENCE/DOCKET NUMBER: GI 5190
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (617) 876-1170
 TELEFAX: (617) 876-5851
 INFORMATION FOR SEQ ID NO: 104:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 1140 amino acids
 TYPE: amino acid
 TOPOLOGY: linear
 MOLECULE TYPE: protein
 SEQUENCE DESCRIPTION: SEQ ID NO: 104:
 US-10-124-557-104

Query Match 99.8%; Score 5872.8; DB 13; Length 1140;
 Best Local Similarity 96.3%; Pred. No. 1.3e-144;
 Matches 1098; Conservative 0; Mismatches 0; Indels 42; Gaps 1;

| | | | |
|----|-----|--|-----|
| QY | 1 | MAWKTLPIVILLLLSVFVIQQVSSODLSSCAGRCGEGYSDATCNCYNCOHYMECCPDF | 60 |
| DB | 1 | MAWKTLPIVILLLLSVFVIQQVSSODLSSCAGRCGEGYSDATCNCYNCOHYMECCPDF | 60 |
| QY | 61 | KRVCTAELSCKRCFESFERGECDCDAQCKKYDKCCPDYSCAEVHNPTSPSSKKAP | 120 |
| DB | 61 | KRVCTAELSCKRCFESFERGECDCDAQCKKYDKCCPDYSCAEVHNPTSPSSKKAP | 120 |
| QY | 121 | PPSGASQITKSTTKSPKPPNKKTKVIESEITE----- | 156 |
| DB | 121 | PPSGASQITKSTTKSPKPPNKKTKVIESEITE----- | 156 |
| QY | 157 | -----KVQDKNKNRTKKXPTKPPVVDAGSLGDLNGDFKVTTPDTST | 198 |
| DB | 181 | KIKSSKNSAANRELQKLVKVDNKNRTKKXPTKPPVVDAGSLGDLNGDFKVTTPDTST | 240 |
| QY | 199 | TOHNKYSTSPKLTAKPINRPSLPNSDTSKETSITVNETKETTNNKQTSIDG | 258 |
| DB | 241 | TOHNKYSTSPKLTAKPINRPSLPNSDTSKETSITVNETKETTNNKQTSIDG | 300 |
| QY | 259 | KEKTTSAKETOSIEKTSADLAPTSKVLAKPTKPAETTKGPAITTKPEPTTPKPEPAS | 318 |
| DB | 301 | KEKTTSAKETOSIEKTSADLAPTSKVLAKPTKPAETTKGPAITTKPEPTTPKPEPAS | 360 |
| QY | 319 | TTPEPTTTIKSAPTTKPEPAPTTTKSAPTTKPEPAPTTTKPEPAPTTTKPEP | 378 |
| DB | 361 | TTPEPTTTIKSAPTTKPEPAPTTTKSAPTTKPEPAPTTTKPEPAPTTTKPEP | 420 |
| QY | 379 | APTTTKSAPTTKPEPAPTTKPKPAPTTKPEPAPTTKPEPAPTTKPEPAPTTK | 438 |
| DB | 421 | APTTTKSAPTTKPEPAPTTKPKPAPTTKPEPAPTTKPEPAPTTKPEPAPTTK | 480 |
| QY | 439 | EPAPTA PKPAPTTKPEPAPTTTKPEPAPTTTKPEPAPTTTKSAPTTTKPEAPT | 498 |
| DB | 481 | EPAPTA PKPAPTTKPEPAPTTTKPEPAPTTTKPEPAPTTTKSAPTTTKPEAPT | 540 |
| QY | 499 | TTKSAPTTKPEPAPTTTKPEPAPTTKPKPAPTTKPEPAPTTKPEPAPTTTKK | 558 |
| DB | 541 | TTKSAPTTKPEPAPTTTKPEPAPTTKPKPAPTTKPEPAPTTKPEPAPTTTKK | 600 |
| QY | 559 | APTAPKPEPAPTTKPEPAPTTTKPELAPTTPEKAPTTPEELAPTTPEEPTPTT | 618 |
| DB | 601 | APTAPKPEPAPTTKPEPAPTTTKPELAPTTPEKAPTTPEELAPTTPEEPTPTT | 660 |
| QY | 619 | PEEPAPTTKAAAANPTKPEPAPTTKPEPAPTTKPEPAPTTKPEPAPTTKPE | 678 |
| DB | 661 | PEEPAPTTKAAAANPTKPEPAPTTKPEPAPTTKPEPAPTTKPEPAPTTKPE | 720 |
| QY | 679 | APTTPKPEPAPKELAPTTTKETSTTSKDPAPTTPKGTAPTTKPEPAPTTKPE | 738 |
| DB | 721 | APTTPKPEPAPKELAPTTTKETSTTSKDPAPTTPKGTAPTTKPEPAPTTKPE | 780 |
| QY | 739 | TAPTTLKPEPAPTTPKKAPKELAPTTTKGTSTTSKDPAPTTKPEPAPTTKPE | 798 |

| | | | |
|----|------|---|------|
| DB | 781 | TAPTTLKPEPAPTTPKKAPKELAPTTTKGTSTTSKDPAPTTKPEPAPTTKPE | 840 |
| QY | 799 | KPAPTTPETPPPTTSVSTPTTKETPTTIHKSPDSTPELSAETPKALENSPKPGVPT | 858 |
| DB | 841 | KPAPTTPETPPPTTSVSTPTTKETPTTIHKSPDSTPELSAETPKALENSPKPGVPT | 900 |
| QY | 859 | TKTPRAATKPEMTTAKDKTTERDLRTPTTTAAAPKWKETATTTKETSITATTTQV | 918 |
| DB | 901 | TKTPRAATKPEMTTAKDKTTERDLRTPTTTAAAPKWKETATTTKETSITATTTQV | 960 |
| QY | 919 | TSTTTQDTTTPFKITLTKTTTLAPKVTTKTITTTTINMKPEETAKPKDRATNSKATTPK | 978 |
| DB | 961 | TSTTTQDTTTPFKITLTKTTTLAPKVTTKTITTTTINMKPEETAKPKDRATNSKATTPK | 1020 |
| QY | 979 | PQKPTKAPKPTSTKPKTNPVRKPKTPTPRKMTSTMPELNPTSRIAEAMLOTTTTPN | 1038 |
| DB | 1021 | PQKPTKAPKPTSTKPKTNPVRKPKTPTPRKMTSTMPELNPTSRIAEAMLOTTTTPN | 1080 |
| QY | 1039 | QTPNSKLVNPKSEBAGAEGETPHMLLRPHVFMPEVTPDMDYLPVFNQGIINPMLS | 1098 |
| DB | 1081 | QTPNSKLVNPKSEBAGAEGETPHMLLRPHVFMPEVTPDMDYLPVFNQGIINPMLS | 1140 |

RESULT 2
 US-09-802-207-30
 ; Sequence 30, Application US/09802207
 ; Publication No. US20020086824A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Warman, Matthew
 ; APPLICANT: Carpten, John
 ; APPLICANT: Trent, Jeffrey
 ; APPLICANT: Marcelino, Jose
 ; TITLE OF INVENTION: Novel Methods and Reagents for the Treatment of Osteoarthritis
 ; FILE REFERENCE: Case-06212
 ; CURRENT APPLICATION NUMBER: US/09/802,207
 ; CURRENT FILING DATE: 2001-08-29
 ; PRIOR APPLICATION NUMBER: 09/619,175
 ; PRIOR FILING DATE: 2000-07-19
 ; PRIOR APPLICATION NUMBER: 60/145,328
 ; PRIOR FILING DATE: 1999-07-23
 ; NUMBER OF SEQ ID NOS: 30
 ; SOFTWARE: Patent in version 3.0
 ; SEQ ID NO 30
 ; LENGTH: 1404
 ; TYPE: PRT
 ; ORGANISM: Homo sapiens
 US-09-802-207-30

Query Match 99.8%; Score 5872.8; DB 9; Length 1404;
 Best Local Similarity 96.3%; Pred. No. 1.6e-144;
 Matches 1098; Conservative 0; Mismatches 0; Indels 42; Gaps 1;

| | | | |
|----|-----|--|-----|
| QY | 1 | MAWKTLPIVILLLLSVFVIQQVSSODLSSCAGRCGEGYSDATCNCYNCOHYMECCPDF | 60 |
| DB | 1 | MAWKTLPIVILLLLSVFVIQQVSSODLSSCAGRCGEGYSDATCNCYNCOHYMECCPDF | 60 |
| QY | 61 | KRVCTAELSCKRCFESFERGECDCDAQCKKYDKCCPDYSCAEVHNPTSPSSKKAP | 120 |
| DB | 61 | KRVCTAELSCKRCFESFERGECDCDAQCKKYDKCCPDYSCAEVHNPTSPSSKKAP | 120 |
| QY | 121 | PPSGASQITKSTTKSPKPPNKKTKVIESEITE----- | 156 |
| DB | 121 | PPSGASQITKSTTKSPKPPNKKTKVIESEITE----- | 156 |
| QY | 157 | -----KVQDKNKNRTKKXPTKPPVVDAGSLGDLNGDFKVTTPDTST | 198 |
| DB | 181 | KIKSSKNSAANRELQKLVKVDNKNRTKKXPTKPPVVDAGSLGDLNGDFKVTTPDTST | 240 |
| QY | 199 | TOHNKYSTSPKLTAKPINRPSLPNSDTSKETSITVNETKETTNNKQTSIDG | 258 |
| DB | 241 | TOHNKYSTSPKLTAKPINRPSLPNSDTSKETSITVNETKETTNNKQTSIDG | 300 |

QY 259 KEKTTSAKETQSIKTSADLAPTSKVLAKPTPKAEITTTKGPALTTPKPTPTTKKPAS 318
Db 301 KEKTTSAKETQSIKTSADLAPTSKVLAKPTPKAEITTTKGPALTTPKPTPTTKKPAS 360
QY 319 TTPKEPTTTIKSAPTTKKEPAPTTTKSAPTTKKEPAPTTTKKAPTTTKKEPAPTTTKKEP 378
Db 361 TTPKEPTTTIKSAPTTKKEPAPTTTKSAPTTKKEPAPTTTKKAPTTTKKEPAPTTTKKEP 420
QY 379 APTTTKSAPTTTKKEPAPTTTKKAPTTTKKEPAPTTTKKEPAPTTTKKEPAPTTTKKEP 438
Db 421 APTTTKSAPTTTKKEPAPTTTKKAPTTTKKEPAPTTTKKEPAPTTTKKEPAPTTTKKEP 480
QY 439 BPAPTAPKAPADTTKKEPAPTTTKKAPTTTKKEPAPTTTKKEPAPTTTKKEPAPTTTKKEP 498
Db 481 BPAPTAPKAPADTTKKEPAPTTTKKAPTTTKKEPAPTTTKKEPAPTTTKKEPAPTTTKKEP 540
QY 499 TTKSAPTTTKKEPAPTTTKKAPTTTKKAPTTTKKAPTTTKKAPTTTKKAPTTTKKAPTTTKK 558
Db 541 TTKSAPTTTKKEPAPTTTKKAPTTTKKAPTTTKKAPTTTKKAPTTTKKAPTTTKKAPTTTKK 600
QY 559 APTAPKAPADTTKKEPAPTTTKKAPTTTKKAPTTTKKAPTTTKKAPTTTKKAPTTTKKAP 618
Db 601 APTAPKAPADTTKKEPAPTTTKKAPTTTKKAPTTTKKAPTTTKKAPTTTKKAPTTTKKAP 660
QY 619 BPAPTAPKAPADTTKKEPAPTTTKKAPTTTKKAPTTTKKAPTTTKKAPTTTKKAPTTTKKAP 678
Db 661 BPAPTAPKAPADTTKKEPAPTTTKKAPTTTKKAPTTTKKAPTTTKKAPTTTKKAPTTTKKAP 720
QY 679 APTTPKAPKELAPTTTKKEPAPTTTKKAPTTTKKAPTTTKKAPTTTKKAPTTTKKAPTTTK 738
Db 721 APTTPKAPKELAPTTTKKEPAPTTTKKAPTTTKKAPTTTKKAPTTTKKAPTTTKKAPTTTK 780
QY 739 TAPTTLKAPADTTKKEPAPTTTKKAPTTTKKAPTTTKKAPTTTKKAPTTTKKAPTTTKKAP 798
Db 781 TAPTTLKAPADTTKKEPAPTTTKKAPTTTKKAPTTTKKAPTTTKKAPTTTKKAPTTTKKAP 840
QY 799 KPAPTTPPTPTTSVSTPTTKKEPAPTTTKKAPTTTKKAPTTTKKAPTTTKKAPTTTKKAP 858
Db 841 KPAPTTPPTPTTSVSTPTTKKEPAPTTTKKAPTTTKKAPTTTKKAPTTTKKAPTTTKKAP 900
QY 859 TKTPAAKPEMTTAKDXTTERDLRTTPTTAAAPKMTKETATTTKTESKITATTQV 918
Db 901 TKTPAAKPEMTTAKDXTTERDLRTTPTTAAAPKMTKETATTTKTESKITATTQV 960
QY 919 TSPTTODTTPFKITTLTKTTLAPKVTITTKITTTIEMNKPBTETAKPKDRATNSKATTPK 978
Db 961 TSPTTODTTPFKITTLTKTTLAPKVTITTKITTTIEMNKPBTETAKPKDRATNSKATTPK 1020
QY 979 PQKPTKAPKPTSTTKKPTMPVRKPTTTPPKMTSTMPNLNPTSRIAEAMLQTTTRPN 1038
Db 1021 PQKPTKAPKPTSTTKKPTMPVRKPTTTPPKMTSTMPNLNPTSRIAEAMLQTTTRPN 1080
QY 1039 QTPNSKLVEVNPKSEBAGGAGETPHMLLRPHVFMPEVTPDMDYLRVFNQGIINPMLS 1098
Db 1081 QTPNSKLVEVNPKSEBAGGAGETPHMLLRPHVFMPEVTPDMDYLRVFNQGIINPMLS 1140

RESULT 3
US-09-897-188-1
; Sequence 1, Application US/09897188
; Publication No. US20040072741A1
; GENERAL INFORMATION:
; APPLICANT: Jay, Gregory D.
; TITLE OF INVENTION: Tribonectin Polypeptides and Uses Thereof
; FILE REFERENCE: 21486-026 CIP2
; CURRENT APPLICATION NUMBER: US/09/897,188
; CURRENT FILING DATE: 2001-07-02
; PRIOR APPLICATION NUMBER: 09/298,970
; PRIOR FILING DATE: 1999-04-23
; PRIOR APPLICATION NUMBER: 09/556,246
; PRIOR FILING DATE: 2000-04-24
; NUMBER OF SEQ ID NOS: 34
; SOFTWARE: Patent In Ver. 2.1

; SEQ ID NO 1
; LENGTH: 1404
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-897-188-1

Query Match 99.8%; Score 5872.8; DB 11; Length 1404;
Best Local Similarity 96.3%; Pred. No. 1.6e-144;
Matches 1098; Conservative 0; Mismatches 0; Indels 42; Gaps 1;

QY 1 MAWKTLPIYLLLLSVFIQQVSSQDLSSCAGCGEGYGRDATCNCDYNCOHYNECCPDF 60
Db 1 MAWKTLPIYLLLLSVFIQQVSSQDLSSCAGCGEGYGRDATCNCDYNCOHYNECCPDF 60
QY 61 KRVTABLSCKGRCFSPFERGRCDDCAQCKKYDKCCPDYESFCAEVHNPTSPSSKKAP 120
Db 61 KRVTABLSCKGRCFSPFERGRCDDCAQCKKYDKCCPDYESFCAEVHNPTSPSSKKAP 120
QY 121 PPSGASOTIKSTTKRSPKPNKKTKKVIBSEBITE----- 156
Db 121 PPSGASOTIKSTTKRSPKPNKKTKKVIBSEBITE----- 180
QY 157 -----KVKNKONTKKKPTPKPPVDEAGSLDNGDFKVTTPDTST 198
Db 181 KIKSSKNSAANRELQKKLVKDNKQRTKKKPTPKPPVDEAGSLDNGDFKVTTPDTST 240
QY 199 TOHNVKSTSPKITTAKFIPNPSLPNSDTSKETSITVANKETTVEKETTITNKQISTDG 258
Db 241 TOHNVKSTSPKITTAKFIPNPSLPNSDTSKETSITVANKETTVEKETTITNKQISTDG 300
QY 259 KEKTTSAKETQSIKTSADLAPTSKVLAKPTPKAEITTTKGPALTTPKPTPTTKKPAS 318
Db 301 KEKTTSAKETQSIKTSADLAPTSKVLAKPTPKAEITTTKGPALTTPKPTPTTKKPAS 360
QY 319 TTPKEPTTTIKSAPTTKKEPAPTTTKSAPTTKKEPAPTTTKKAPTTTKKEPAPTTTKKEP 378
Db 361 TTPKEPTTTIKSAPTTKKEPAPTTTKSAPTTKKEPAPTTTKKAPTTTKKEPAPTTTKKEP 420
QY 379 APTTTKSAPTTTKKEPAPTTTKKAPTTTKKAPTTTKKAPTTTKKAPTTTKKAPTTTKKAP 438
Db 421 APTTTKSAPTTTKKEPAPTTTKKAPTTTKKAPTTTKKAPTTTKKAPTTTKKAPTTTKKAP 480
QY 439 BPAPTAPKAPADTTKKEPAPTTTKKAPTTTKKAPTTTKKAPTTTKKAPTTTKKAPTTTKKAP 498
Db 481 BPAPTAPKAPADTTKKEPAPTTTKKAPTTTKKAPTTTKKAPTTTKKAPTTTKKAPTTTKKAP 540
QY 499 TTKSAPTTTKKEPAPTTTKKAPTTTKKAPTTTKKAPTTTKKAPTTTKKAPTTTKKAPTTTKK 558
Db 541 TTKSAPTTTKKEPAPTTTKKAPTTTKKAPTTTKKAPTTTKKAPTTTKKAPTTTKKAPTTTKK 600
QY 559 APTAPKAPADTTKKEPAPTTTKKAPTTTKKAPTTTKKAPTTTKKAPTTTKKAPTTTKKAP 618
Db 601 APTAPKAPADTTKKEPAPTTTKKAPTTTKKAPTTTKKAPTTTKKAPTTTKKAPTTTKKAP 660
QY 619 BPAPTAPKAPADTTKKEPAPTTTKKAPTTTKKAPTTTKKAPTTTKKAPTTTKKAPTTTKKAP 678
Db 661 BPAPTAPKAPADTTKKEPAPTTTKKAPTTTKKAPTTTKKAPTTTKKAPTTTKKAPTTTKKAP 720
QY 679 APTTPKAPKELAPTTTKKEPAPTTTKKAPTTTKKAPTTTKKAPTTTKKAPTTTKKAPTTTK 738
Db 721 APTTPKAPKELAPTTTKKEPAPTTTKKAPTTTKKAPTTTKKAPTTTKKAPTTTKKAPTTTK 780
QY 739 TAPTTLKAPADTTKKEPAPTTTKKAPTTTKKAPTTTKKAPTTTKKAPTTTKKAPTTTKKAP 798
Db 781 TAPTTLKAPADTTKKEPAPTTTKKAPTTTKKAPTTTKKAPTTTKKAPTTTKKAPTTTKKAP 840
QY 799 KPAPTTPPTPTTSVSTPTTKKEPAPTTTKKAPTTTKKAPTTTKKAPTTTKKAPTTTKKAP 858
Db 841 KPAPTTPPTPTTSVSTPTTKKEPAPTTTKKAPTTTKKAPTTTKKAPTTTKKAPTTTKKAP 900
QY 859 TKTPAAKPEMTTAKDXTTERDLRTTPTTAAAPKMTKETATTTKTESKITATTQV 918
Db 901 TKTPAAKPEMTTAKDXTTERDLRTTPTTAAAPKMTKETATTTKTESKITATTQV 960

RESULT 5
US-10-124-557-62
; Sequence 62, Application US/10124557
; Publication No. US20020137894A1
; GENERAL INFORMATION:
; APPLICANT: Turner, Katherine
; Clark, Stephen C.
; Jacobs, Kenneth
; Hewick, Rodney M.
; Gesner, Thomas G.
; TITLE OF INVENTION: Megakaryocyte Stimulating Factors
; NUMBER OF SEQUENCES: 143
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genetics Institute, Inc.
; STREET: 87 CambridgePark Drive
; CITY: Cambridge
; STATE: Massachusetts
; COUNTRY: U.S.A.
; ZIP: 02140
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent in Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/10/124,557
; FILING DATE: 16-Apr-2002
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/643,502
; FILING DATE: 18-JAN-1991
; APPLICATION NUMBER: US 07/546,114
; FILING DATE: 29-JUN-1990
; APPLICATION NUMBER: US 07/457,196
; FILING DATE: 29-DEC-1989
; APPLICATION NUMBER: US 07/390,901
; FILING DATE: 08-AUG-1989
; ATTORNEY/AGENT INFORMATION:
; NAME: Cseri, Luann
; REGISTRATION NUMBER: 31,822
; REFERENCE/DOCKET NUMBER: G1 5190
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617)876-1170
; TELEFAX: (617)876-5851
; INFORMATION FOR SEQ ID NO: 62:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1404 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; SEQUENCE DESCRIPTION: SEQ ID NO: 62:
US-10-124-557-62
Query Match 99.8%; Score 5872.8; DB 13; Length 1404;
Best Local Similarity 96.3%; Pred. No. 1.6e-144;
Matches 1098; Conservative 0; Mismatches 0; Indels 42; Gaps 1;
QY 1 MANKTPIYLLLSVPIQVSSQDLSSCAGCGGYSRDATAICNDYNCQHYNECCPDF 60
DB 1 MANKTPIYLLLSVPIQVSSQDLSSCAGCGGYSRDATAICNDYNCQHYNECCPDF 60
QY 61 KRVCATLSCKGRCFESFERGECDDAQCKKYDKCCPDYESFCAEVHNPTSPSSSKAP 120
DB 61 KRVCATLSCKGRCFESFERGECDDAQCKKYDKCCPDYESFCAEVHNPTSPSSSKAP 120
QY 121 PPAGASQTIKSTTKRSPKPNKKTKVIESEITE----- 156
DB 121 PPAGASQTIKSTTKRSPKPNKKTKVIESEITE----- 180
QY 157 -----KVKDNKNRTKKPKPPVVDVDEAGSLDNGDFKVTTPDTST 198

RESULT 6

US-10-124-557-40
; Sequence 40, Application US/10124557
; Publication No. US20020137894A1
; GENERAL INFORMATION:
; APPLICANT: Turner, Katherine
; Clark, Stephen C.
; Jacobs, Kenneth
; Hewick, Rodney M.
; Gesner, Thomas G.

TITLE OF INVENTION: Megakaryocyte Stimulating Factors
 NUMBER OF SEQUENCES: 143
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: Genetics Institute, Inc.
 STREET: 87 CambridgePark Drive
 CITY: Cambridge
 STATE: Massachusetts
 COUNTRY: U.S.A.
 ZIP: 02140
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: Patentin Release #1.0, Version #1.25
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/10/124,557
 FILING DATE: 16-Apr-2002
 CLASSIFICATION: <Unknown>
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 07/643,502
 FILING DATE: 18-JAN-1991
 APPLICATION NUMBER: US 07/546,114
 FILING DATE: 29-JUN-1990
 APPLICATION NUMBER: US 07/457,196
 FILING DATE: 29-DEC-1989
 APPLICATION NUMBER: US 07/390,901
 FILING DATE: 08-AUG-1989
 ATTORNEY/AGENT INFORMATION:
 NAME: Cseir, Luann
 REGISTRATION NUMBER: 31,822
 REFERENCE/DOCKET NUMBER: GI 5190
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (617)876-1170
 TELEFAX: (617)876-5851
 INFORMATION FOR SEQ ID NO: 40:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 1361 amino acids
 TYPE: amino acid
 TOPOLOGY: linear
 MOLECULE TYPE: protein
 SEQUENCE DESCRIPTION: SEQ ID NO: 40:

US-10-124-557-40
 Query Match 99.7%; Score 5871.9; DB 13; Length 1361;
 Best Local Similarity 99.9%; Pred. No. 1.6e-144;
 Matches 1097; Conservative 0; Mismatches 0; Indels 1; Gaps 1;
 QY 1 MAWKTLPIYLILLISVFIQVSSQDLSSCAGRCGEYSRDTNCNDYNCQHYMECCPDF 60
 DB 1 MAWKTLPIYLILLISVFIQVSSQDLSSCAGRCGEYSRDTNCNDYNCQHYMECCPDF 60
 QY 61 KRVTAEISCKGRCFESPERGECDDAQCKYDKCCPDYFSCAEVHNPTSPSSKKAP 120
 DB 61 KRVTAEISCKGRCFESPERGECDDAQCKYDKCCPDYFSCAEVHNPTSPSSKKAP 120
 QY 121 PPSGASQTIKSTTKRSPKPNKKTKKVISEEITEKVNKKQNRKTKKTPKPPVVDAA 180
 DB 121 PPSGASQTIKSTTKRSPKPNKKTKKVISEEITE-KVNKKQNRKTKKTPKPPVVDAA 179
 QY 181 GSGLDNGDFKVTTPDTSTTQHNKYSTSPKLTAKPINPRESLPNSDTSKETSITVKNK 240
 DB 180 GSGLDNGDFKVTTPDTSTTQHNKYSTSPKLTAKPINPRESLPNSDTSKETSITVKNK 239
 QY 241 TVETKETTNNKQSTQDKETTSKETSQTSKETSADKLAPTSQVLAETPKAETTKGP 300
 DB 240 TVETKETTNNKQSTQDKETTSKETSQTSKETSADKLAPTSQVLAETPKAETTKGP 299
 QY 301 ALTTPEKPTPTTKPEASTTKPEPTPTTKSAPTTKEPAPTTTKSAPTTKEPAPTTTK 360
 DB 300 ALTTPEKPTPTTKPEASTTKPEPTPTTKSAPTTKEPAPTTTKSAPTTKEPAPTTTK 359
 QY 361 EPAPTTKEPAPTTTKPEAPTTTKSAPTTKEPAPTTTKPKKAPTTKEPAPTTKEPAPTT 420

DB 360 EPAPTTKEPAPTTTKPEAPTTTKSAPTTKEPAPTTTKPKKAPTTKEPAPTTKEPAPTT 419
 QY 421 TPKEPAPTTKEPAPTTTKPEAPTTAPKAPADTTKEPAPTTKEPAPTTKEPAPTTKEPAPTT 480
 DB 420 TPKEPAPTTKEPAPTTTKPEAPTTAPKAPADTTKEPAPTTKEPAPTTKEPAPTTKEPAPTT 479
 QY 481 APITTSAPTTTKPEAPTTTKSAPTTTKPEAPTTTKPEAPTTTKPEAPTTTKPEAPTTTK 540
 DB 480 APITTSAPTTTKPEAPTTTKSAPTTTKPEAPTTTKPEAPTTTKPEAPTTTKPEAPTTTK 539
 QY 541 EPAPTTKEPAPTTTKKAPADTTKEPAPTTKEPAPTTTKKAPADTTKEPAPTTTKKAPADTT 600
 DB 540 EPAPTTKEPAPTTTKKAPADTTKEPAPTTTKKAPADTTKEPAPTTTKKAPADTTKEPAPTT 599
 QY 601 TTPEELAPTTPEBPTPTTPEEPAPTTPKAAAPNTPEEPAPTTPKAAAPNTPEEPAPTTPK 660
 DB 600 TTPEELAPTTPEBPTPTTPEEPAPTTPKAAAPNTPEEPAPTTPKAAAPNTPEEPAPTTPK 659
 QY 661 ETAPTTKGTAPTTLKEPAPTTPKKAPKELAPTTTKEPTSTTSKAPADTTTKGTAPTTT 720
 DB 660 ETAPTTKGTAPTTLKEPAPTTPKKAPKELAPTTTKEPTSTTSKAPADTTTKGTAPTTT 719
 QY 721 KEAPTTKEPAPTTTKGTAPTTLKEPAPTTPKKAPKELAPTTTKGTAPTTSDKAPTT 780
 DB 720 KEAPTTKEPAPTTTKGTAPTTLKEPAPTTPKKAPKELAPTTTKGTAPTTSDKAPTT 779
 QY 781 PKETAPTTKEPAPTTPKKAPADTTTETPPPTTSEVSTPTTKEPTTIHKSPDESTPELSA 840
 DB 780 PKETAPTTKEPAPTTPKKAPADTTTETPPPTTSEVSTPTTKEPTTIHKSPDESTPELSA 839
 QY 841 EPTPKALENSPKPEGVPTTKTPAAATKPEMTTAKOKTTERDLRTTETTTAAAPKMTKETA 900
 DB 840 EPTPKALENSPKPEGVPTTKTPAAATKPEMTTAKOKTTERDLRTTETTTAAAPKMTKETA 899
 QY 901 TTTEKTTESKITATTTQVSTTTTQDTTPFKITLTKTTTLAPKVTTKTKTTTTEIMNKPE 960
 DB 900 TTTEKTTESKITATTTQVSTTTTQDTTPFKITLTKTTTLAPKVTTKTKTTTTEIMNKPE 959
 QY 961 ETAKPKDRATNSKATTPKPKQKTPKAPKPTSTKPKKPTWVRVKKPTTTPRKMSTMP 1020
 DB 960 ETAKPKDRATNSKATTPKPKQKTPKAPKPTSTKPKKPTWVRVKKPTTTPRKMSTMP 1019
 QY 1021 NPTSRIAEAMLOTTTRPNQTPNSKLVEVNPKSEDAAGAEGETPHMLLRPHVMPVETPDM 1080
 DB 1020 NPTSRIAEAMLOTTTRPNQTPNSKLVEVNPKSEDAAGAEGETPHMLLRPHVMPVETPDM 1079
 QY 1081 DYLPRVNPQGIINPMLS 1098
 DB 1080 DYLPRVNPQGIINPMLS 1097

RESULT 7

US-10-124-557-48
 ; Sequence 48, Application US/10124557
 ; Publication No. US20020137894A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Turner, Katherine
 ; Jacobs, Kenneth
 ; Hewick, Rodney M.
 ; Gesner, Thomas G.
 ; TITLE OF INVENTION: Megakaryocyte Stimulating Factors
 ; NUMBER OF SEQUENCES: 143
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Genetics Institute, Inc.
 ; STREET: 87 CambridgePark Drive
 ; CITY: Cambridge
 ; STATE: Massachusetts
 ; COUNTRY: U.S.A.
 ; ZIP: 02140
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible


```

; FILING DATE: 29-DEC-1989
; APPLICATION NUMBER: US 07/390,901
; FILING DATE: 09-AUG-1989
; ATTORNEY/AGENT INFORMATION:
;   NAME: Cserr, Luann
;   REGISTRATION NUMBER: 31,822
;   REFERENCE/DOCKET NUMBER: GI 5190
;   TELECOMMUNICATION INFORMATION:
;     TELEPHONE: (617)876-1170
;     TELEFAX: (617)876-5851
; INFORMATION FOR SEQ ID NO: 52:
; SEQUENCE CHARACTERISTICS:
;   LENGTH: 1363 amino acids
;   TYPE: amino acid
;   TOPOLOGY: linear
;   MOLECULE TYPE: protein
; SEQUENCE DESCRIPTION: SEQ ID NO: 52:
US-10-124-557-52

Query Match      95.3%; Score 5608.7; DB 13; Length 1363;
Best Local Similarity 92.7%; Pred. No. 1.1e-137;
Matches 1057; Conservative 0; Mismatches 0; Indels 83; Gaps 2;

QY 1 MAWKTLPIYLLLLSVFVIQVSSODLSACAGCGEGYSRDATCNCDYNCQHYMECCPDF 60
DB 1 MAWKTLPIYLLLLSVFVIQVSSQ----- 25

QY 61 KRVTCTAELSCGRCPESFERGECDCDAQCKKYDKCCPDYBSCAEVHNPTSPSSKKAP 120
DB 26 -----ELSCGRCPESFERGECDCDAQCKKYDKCCPDYBSCAEVHNPTSPSSKKAP 79

QY 121 PPSGASQIKSTKSPRPNNKKTKVIEBEEITE----- 156
DB 80 PPSGASQIKSTKSPRPNNKKTKVIEBEEITE----- 139

QY 157 -----KVQDKNKNRTKKPTKPPVVDVAGSLDNGDFKVTPTST 198
DB 140 KIKSSNKAANRELQKULKVDKNKNRTKKPTKPPVVDVAGSLDNGDFKVTPTST 199

QY 199 TQHNKYSTPKITTAAPINRPSLPNSDTSKESLTVNKTETVETKTTTTNNQSTDG 258
DB 200 TQHNKYSTPKITTAAPINRPSLPNSDTSKESLTVNKTETVETKTTTTNNQSTDG 259

QY 259 KEKTTSAKETQSIKTSKADLAFTSKVLAKPTKAEATTKGPAITTKPEPTTPPKSPAS 318
DB 260 KEKTTSAKETQSIKTSKADLAFTSKVLAKPTKAEATTKGPAITTKPEPTTPPKSPAS 319

QY 319 TTPKEPTTTIKSAPTTKPEAPTTTKSAPTTKPEAPTTTKPEAPTTTKPEAPTTTKPE 378
DB 320 TTPKEPTTTIKSAPTTKPEAPTTTKSAPTTKPEAPTTTKPEAPTTTKPEAPTTTKPE 379

QY 379 ABTTTKSAPTTKPEAPTTKPEAPTTKPEAPTTKPEAPTTKPEPTTTPKPEAPTTKPE 438
DB 380 ABTTTKSAPTTKPEAPTTKPEAPTTKPEAPTTKPEAPTTKPEPTTTPKPEAPTTKPE 439

QY 439 EPAPTAPKKPAPTTPKPEAPTTTPKPEAPTTTKKESPTTPKPEAPTTTKSAPTTTKPE 498
DB 440 EPAPTAPKKPAPTTPKPEAPTTTPKPEAPTTTKKESPTTPKPEAPTTTKSAPTTTKPE 499

QY 499 TTKSAPTTTKKESPTTPKPEAPTTTPKPEAPTTTPKPEAPTTTPKPEAPTTTKPKP 558
DB 500 TTKSAPTTTKKESPTTPKPEAPTTTPKPEAPTTTPKPEAPTTTPKPEAPTTTKPKP 559

QY 559 APTAPKPEAPTTKPEAPTTTPKPEAPTTTPKPEAPTTTPKPEAPTTPEELAPTTPEPT 618
DB 560 APTAPKPEAPTTKPEAPTTTPKPEAPTTTPKPEAPTTTPKPEAPTTPEELAPTTPEPT 619

QY 619 PEEAPTTPKAAAPNTPKPEAPTTTPKPEAPTTTPKPEAPTTTPKPEAPTTTPKGTAT 678
DB 620 PEEAPTTPKAAAPNTPKPEAPTTTPKPEAPTTTPKPEAPTTTPKPEAPTTTPKGTAT 679

QY 679 APTTPKKAPKELAPTTTKETSTSDKPAITTPKGTAPTTTPKPEAPTTTPKPEAPTTPKG 738

```

```

DB 680 APTTPKKAPKELAPTTTKETSTSDKPAITTPKGTAPTTTPKPEAPTTTPKPEAPTTPKG 739
QY 739 TAPTTLLKEPAPTTPKKAPKELAPTTTKGTPTSTSDKPAITTPKGTAPTTTPKPEAPTTPK 798
DB 740 TAPTTLLKEPAPTTPKKAPKELAPTTTKGTPTSTSDKPAITTPKGTAPTTTPKPEAPTTPK 799
QY 799 KPAPTTPEPTTSEVSTPTTTKPTTTHKSPDSESTPELSAETPKALENSPKSPGVPT 858
DB 800 KPAPTTPEPTTSEVSTPTTTKPTTTHKSPDSESTPELSAETPKALENSPKSPGVPT 859
QY 859 TKTPAATKPEMTTAXDKTTERDLRTPTTTPETTTTAAAPKMTKETATTTKTTESKITATTQV 918
DB 860 TKTPAATKPEMTTAXDKTTERDLRTPTTTPETTTTAAAPKMTKETATTTKTTESKITATTQV 919
QY 919 TSTTTQDTTTPFKITTLKTTTLAPKVTITTKTITTTTINMKPEETAKPKDRATNSKATTPK 978
DB 920 TSTTTQDTTTPFKITTLKTTTLAPKVTITTKTITTTTINMKPEETAKPKDRATNSKATTPK 979
QY 979 PQKPTKAPKPTSTKPKTMPRVKPKTTPPKKVTSTMPKVTSTMPKVTSTMPKVTSTMPKVTST 1038
DB 980 PQKPTKAPKPTSTKPKTMPRVKPKTTPPKKVTSTMPKVTSTMPKVTSTMPKVTSTMPKVTST 1039
QY 1039 QTPNSKLVNPKSEADAGAEGETPHMLLRPHVFMPEVTPDMDYLPRVFNQGIINPMLS 1098
DB 1040 QTPNSKLVNPKSEADAGAEGETPHMLLRPHVFMPEVTPDMDYLPRVFNQGIINPMLS 1099

RESULT 9
US-10-124-557-42
; Sequence 42, Application US/10124557
; Publication No. US20020137894A1
; GENERAL INFORMATION:
; APPLICANT: Turner, Katherine
; Clark, Stephen C.
; Jacobs, Kenneth
; Hewick, Rodney M.
; Gesner, Thomas G.
; TITLE OF INVENTION: Megakaryocyte Stimulating Factors
; NUMBER OF SEQUENCES: 143
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genetics Institute, Inc.
; STREET: 87 CambridgePark Drive
; City: Cambridge
; STATE: Massachusetts
; COUNTRY: U.S.A.
; ZIP: 02140
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA: US/10/124,557
; FILING DATE: 16-Apr-2002
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/643,502
; FILING DATE: 18-JAN-1991
; APPLICATION NUMBER: US 07/546,114
; FILING DATE: 29-JUN-1990
; APPLICATION NUMBER: US 07/457,196
; FILING DATE: 29-DEC-1989
; APPLICATION NUMBER: US 07/390,901
; FILING DATE: 08-AUG-1989
; ATTORNEY/AGENT INFORMATION:
; NAME: Cserr, Luann
; REGISTRATION NUMBER: 31,822
; REFERENCE/DOCKET NUMBER: GI 5190
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617)876-1170
; TELEFAX: (617)876-5851
; INFORMATION FOR SEQ ID NO: 42:
; SEQUENCE CHARACTERISTICS:

```

```
; ; LENGTH: 1311 amino acids
; ; TYPE: amino acid
; ; TOPOLOGY: linear
; ; MOLECULE TYPE: protein
; ; SEQUENCE DESCRIPTION: SEQ ID NO: 42:
US-10-124-557-42

Query Match          95.3%; Score 5607.9; DB 13; Length 1311;
Best Local Similarity 95.4%; Pred. No. 1.1e-137;
Matches 1047; Conservative 0; Mismatches 0; Indels 51; Gaps 1;

QY 1 MAWKTLPIYLLLLSVFVYQVSSQDLSSCAGCGEGYSRDATCNCYNQCQHYMECCPDF 60
DB 1 MAWKTLPIYLLLLSVFVYQVSSQDLSSCAGCGEGYSRDATCNCYNQCQHYMECCPDF 60

QY 61 KRVTAEALCKGRCFESFERGECDCDAQCKKYDKCCPDYEFCAEYHNPTSPSPSKXAP 120
DB 61 KRVTAEALCKGRCFESFERGECDCDAQCKKYDKCCPDYEFCAEYHNPTSPSPSKXAP 120

QY 121 PFGASQTIKSTTKRSPKPNKKTKVIESEBIEIKVKDKNKNRTKKKTPKPPVDEA 180
DB 121 PFGASQTIKSTTKRSPKPNKKTKVIESEBIEIKVKDKNKNRTKKKTPKPPVDEA 180

QY 107 -----VKDNKNRTKKKTPKPPVDEA 129
DB 107 -----VKDNKNRTKKKTPKPPVDEA 129

QY 181 GSGLDNGDFKVTTPDTSTTQHNKVSTSPKITTAKPINRPSLPNSDTSKETSIVNKET 240
DB 181 GSGLDNGDFKVTTPDTSTTQHNKVSTSPKITTAKPINRPSLPNSDTSKETSIVNKET 240

QY 130 GSGLDNGDFKVTTPDTSTTQHNKVSTSPKITTAKPINRPSLPNSDTSKETSIVNKET 189
DB 130 GSGLDNGDFKVTTPDTSTTQHNKVSTSPKITTAKPINRPSLPNSDTSKETSIVNKET 189

QY 241 TVETKETTITNTQSTDGKEKTSIAKETOSIEKTSADKLAPTSKVLAKPTPAETTTKGP 300
DB 241 TVETKETTITNTQSTDGKEKTSIAKETOSIEKTSADKLAPTSKVLAKPTPAETTTKGP 300

QY 190 TVETKETTITNTQSTDGKEKTSIAKETOSIEKTSADKLAPTSKVLAKPTPAETTTKGP 249
DB 190 TVETKETTITNTQSTDGKEKTSIAKETOSIEKTSADKLAPTSKVLAKPTPAETTTKGP 249

QY 301 ALTTKPEPTTPKPEASTPKPEPTTTKSAPTTPKEPATTTKSAPTTPKEPATTTK 360
DB 301 ALTTKPEPTTPKPEASTPKPEPTTTKSAPTTPKEPATTTKSAPTTPKEPATTTK 360

QY 250 ALTTKPEPTTPKPEASTPKPEPTTTKSAPTTPKEPATTTKSAPTTPKEPATTTK 309
DB 250 ALTTKPEPTTPKPEASTPKPEPTTTKSAPTTPKEPATTTKSAPTTPKEPATTTK 309

QY 361 EPAPTTKPEPATTTKPEPATTTKSAPTTPKEPATTTKPKAPATTPKBPATTPKEPT 420
DB 361 EPAPTTKPEPATTTKPEPATTTKSAPTTPKEPATTTKPKAPATTPKBPATTPKEPT 420

QY 310 EPAPTTKPEPATTTKPEPATTTKSAPTTPKEPATTTKPKAPATTPKBPATTPKEPT 369
DB 310 EPAPTTKPEPATTTKPEPATTTKSAPTTPKEPATTTKPKAPATTPKBPATTPKEPT 369

QY 421 TPKEPATTTKPEPATTTKPEPATTTKPEPATTTKPEPATTTKPEPATTTKPEPAT 480
DB 421 TPKEPATTTKPEPATTTKPEPATTTKPEPATTTKPEPATTTKPEPATTTKPEPAT 480

QY 370 TPKEPATTTKPEPATTTKPEPATTTKPEPATTTKPEPATTTKPEPATTTKPEPAT 429
DB 370 TPKEPATTTKPEPATTTKPEPATTTKPEPATTTKPEPATTTKPEPATTTKPEPAT 429

QY 481 APTTKSAPTTTKPEPATTTKSAPTTPKESPTTTKPEPATTTKPEPATTTKPEPAT 540
DB 481 APTTKSAPTTTKPEPATTTKSAPTTPKESPTTTKPEPATTTKPEPATTTKPEPAT 540

QY 430 APTTKSAPTTTKPEPATTTKSAPTTPKESPTTTKPEPATTTKPEPATTTKPEPAT 489
DB 430 APTTKSAPTTTKPEPATTTKSAPTTPKESPTTTKPEPATTTKPEPATTTKPEPAT 489

QY 541 EPAPTTKPEPATTTKPEPATTTKPEPATTTKPEPATTTKPEPATTTKPEPATTTK 600
DB 541 EPAPTTKPEPATTTKPEPATTTKPEPATTTKPEPATTTKPEPATTTKPEPATTTK 600

QY 490 EPAPTTKPEPATTTKPEPATTTKPEPATTTKPEPATTTKPEPATTTKPEPATTTK 549
DB 490 EPAPTTKPEPATTTKPEPATTTKPEPATTTKPEPATTTKPEPATTTKPEPATTTK 549

QY 601 TTPPELAPTTPEPTTPPEPATTTPKAAAPNTPEPATTTPEPATTTPEPATTTPK 660
DB 601 TTPPELAPTTPEPTTPPEPATTTPKAAAPNTPEPATTTPEPATTTPEPATTTPK 660

QY 550 TTPPELAPTTPEPTTPPEPATTTPKAAAPNTPEPATTTPEPATTTPEPATTTPK 609
DB 550 TTPPELAPTTPEPTTPPEPATTTPKAAAPNTPEPATTTPEPATTTPEPATTTPK 609

QY 661 ETAPTTKGTATTTKEPATTTKPEPATTTKPEPATTTKPEPATTTKPEPATTTK 720
DB 661 ETAPTTKGTATTTKEPATTTKPEPATTTKPEPATTTKPEPATTTKPEPATTTK 720

QY 610 ETAPTTKGTATTTKEPATTTKPEPATTTKPEPATTTKPEPATTTKPEPATTTK 669
DB 610 ETAPTTKGTATTTKEPATTTKPEPATTTKPEPATTTKPEPATTTKPEPATTTK 669

QY 721 KEPATTTKPEPATTTKGTATTTKEPATTTKPEPATTTKPEPATTTKPEPATTTK 780
DB 721 KEPATTTKPEPATTTKGTATTTKEPATTTKPEPATTTKPEPATTTKPEPATTTK 780

QY 670 KEPATTTKPEPATTTKGTATTTKEPATTTKPEPATTTKPEPATTTKPEPATTTK 729
DB 670 KEPATTTKPEPATTTKGTATTTKEPATTTKPEPATTTKPEPATTTKPEPATTTK 729

QY 781 PKETAPTTKPEPATTTKPEPATTTKPEPATTTKPEPATTTKPEPATTTKPEPAT 840
DB 781 PKETAPTTKPEPATTTKPEPATTTKPEPATTTKPEPATTTKPEPATTTKPEPAT 840

QY 730 PKETAPTTKPEPATTTKPEPATTTKPEPATTTKPEPATTTKPEPATTTKPEPAT 789
DB 730 PKETAPTTKPEPATTTKPEPATTTKPEPATTTKPEPATTTKPEPATTTKPEPAT 789

QY 841 ETPKALENSPKGPGVPTTKTAAPKPEMTTAKDKTTERDLRTTPETTTAAPKMTETA 900
DB 841 ETPKALENSPKGPGVPTTKTAAPKPEMTTAKDKTTERDLRTTPETTTAAPKMTETA 900

QY 790 ETPKALENSPKGPGVPTTKTAAPKPEMTTAKDKTTERDLRTTPETTTAAPKMTETA 849
DB 790 ETPKALENSPKGPGVPTTKTAAPKPEMTTAKDKTTERDLRTTPETTTAAPKMTETA 849

QY 901 TTEKTTESKITATTTQVSTTTQDTPPKITTLKTTTLAPKVTITTKITTTIMNKPE 960
DB 901 TTEKTTESKITATTTQVSTTTQDTPPKITTLKTTTLAPKVTITTKITTTIMNKPE 960
```

Query Match 95.3%; Score 5607.8; DB 13; Length 1320;
Best Local Similarity 96.2%; Pred. No. 1.1e-137;
Matches 1056; Conservative 0; Mismatches 0; Indels 42; Gaps 2;

QY 1 MAWKTLPIYLLLLSVFVYQVSSQDLSSCAGCGEGYSRDATCNCYNQCQHYMECCPDF 60

Db 1 MAWKTLPIYLLLLSVFVIQVSSQ----- 25

QY 61 KRVTCTAELSCGRCEPESFERGECDCDAQCKYDKCCPDYESFCAEVHNPTSPSSKKAP 120

Db 26 -----EUSCKGRCEPESFERGECDCDAQCKYDKCCPDYESFCAEVHNPTSPSSKKAP 79

QY 121 PPSGASQTIKSTTKRSPKPPNKKTKKVIIESEITEKVONKKNRTKKKTKPKPPVVD 180

Db 80 PPSGASQTIKSTTKRSPKPPNKKTKKVIIESEITE -VKONKKNRTKKKTKPKPPVVD 138

QY 181 GSGLDNGDFKVTTPDSTTQHNKYSTSPKITTAKPINRPSLPNSDTSKETSITVNKET 240

Db 139 GSGLDNGDFKVTTPDSTTQHNKYSTSPKITTAKPINRPSLPNSDTSKETSITVNKET 198

QY 241 TVETKETTITNKQSTDCGKETTSAKETQSIKTSKADLAPTSKVLAKPTPKAETTTKGP 300

Db 199 TVETKETTITNKQSTDCGKETTSAKETQSIKTSKADLAPTSKVLAKPTPKAETTTKGP 258

QY 301 ALTTPEKPTPTPKBPASTTPEKPTPTTIKSAPTTPKEPAPTTTKSAPTTPKBPAPTTTK 360

Db 259 ALTTPEKPTPTPKBPASTTPEKPTPTTIKSAPTTPKEPAPTTTKSAPTTPKBPAPTTTK 318

QY 361 EPAPTTPEKAPTTTKBPAPTTTKSAPTTPKBPAPTTPKBPAPTTPKBPAPTTPKBPAPTT 420

Db 319 EPAPTTPEKAPTTTKBPAPTTTKSAPTTPKBPAPTTPKBPAPTTPKBPAPTTPKBPAPTT 378

QY 421 TPKEPAPTTKEPAPTTKEPAPTTAPKBPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTT 480

Db 379 TPKEPAPTTKEPAPTTKEPAPTTAPKBPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTT 438

QY 481 APTTTKSAPTTKEPAPTTTKSAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTT 540

Db 439 APTTTKSAPTTKEPAPTTTKSAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTT 498

QY 541 EPAPTTPEKAPTTTKBPAPTTTKBPAPTTPKBPAPTTPKBPAPTTPKBPAPTTPKBPAPTT 600

Db 499 EPAPTTPEKAPTTTKBPAPTTTKBPAPTTPKBPAPTTPKBPAPTTPKBPAPTTPKBPAPTT 558

QY 601 TTPBELAPTTPEEPTPTTPEEPTPTTPEEPTPTTPEEPTPTTPEEPTPTTPEEPTPTT 660

Db 559 TTPBELAPTTPEEPTPTTPEEPTPTTPEEPTPTTPEEPTPTTPEEPTPTTPEEPTPTT 618

QY 661 ETAPTTTPKGTAPTTLKEPAPTTTPKAPAPKELAPTTTKEPTSTTSDKAPATPKGTATTP 720

Db 619 ETAPTTTPKGTAPTTLKEPAPTTTPKAPAPKELAPTTTKEPTSTTSDKAPATPKGTATTP 678

QY 721 KEAPTTTPKEPAPTTPKGTATPTTLKEPAPTTTPKAPAPKELAPTTTKEPTSTTSDKAPAT 780

Db 679 KEAPTTTPKEPAPTTPKGTATPTTLKEPAPTTTPKAPAPKELAPTTTKEPTSTTSDKAPAT 738

QY 781 PKSTAPTTKEPAPTTTPKAPAPKELAPTTTKEPTSTTSDKAPATPKGTATPKGTATPK 840

Db 739 PKSTAPTTKEPAPTTTPKAPAPKELAPTTTKEPTSTTSDKAPATPKGTATPKGTATPK 798

QY 841 EPTPKALENSKEPGVPTTKTAPATKPEMTTAKDKTTERDLRTTPTTAAAPKWTXETA 900

Db 799 EPTPKALENSKEPGVPTTKTAPATKPEMTTAKDKTTERDLRTTPTTAAAPKWTXETA 858

QY 901 TTTETKTSKATATTTQVSTTQDTPPKITLTKTTLAPKVTTKTITTTIMNKPE 960

Db 859 TTTETKTSKATATTTQVSTTQDTPPKITLTKTTLAPKVTTKTITTTIMNKPE 918

QY 961 ETAKPKDRATNSKATTPKPKQPKTAPKPKTSTTKPKTTPKPKTTPKPKTTPKPKTTPK 1020

Db 919 ETAKPKDRATNSKATTPKPKQPKTAPKPKTSTTKPKTTPKPKTTPKPKTTPKPKTTPK 978

QY 1021 NPTSRIAEAMLOTTTRPNQNTNSKLVNPKSEDAGGAGETPHMLLPHVFMPEVTPDM 1080

Db 979 NPTSRIAEAMLOTTTRPNQNTNSKLVNPKSEDAGGAGETPHMLLPHVFMPEVTPDM 1038

QY 1081 DYLPRVFNQGIILNPMLS 1098

Db 1039 DYLPRVFNQGIILNPMLS 1056

RESULT 11

US-10-124-557-60

; Sequence 60, Application US/10124557

; Publication No. US20020137894A1

; GENERAL INFORMATION:

; APPLICANT: Turner, Katherine

; Clark, Stephen C.

; Jacobs, Kenneth

; Hewick, Rodney M.

; Gesner, Thomas G.

; TITLE OF INVENTION: Megakaryocyte Stimulating Factors

; NUMBER OF SEQUENCES: 143

; CORRESPONDENCE ADDRESS:

; ADDRESS: Genetics Institute, Inc.

; STREET: 87 CambridgePark Drive

; CITY: Cambridge

; STATE: Massachusetts

; COUNTRY: U.S.A.

; ZIP: 02140

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: PatentIn Release #1.0, Version #1.25

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/10/124,557

; FILING DATE: 16-Apr-2002

; CLASSIFICATION: <Unknown>

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: US 07/643,502

; FILING DATE: 18-JAN-1991

; APPLICATION NUMBER: US 07/546,114

; FILING DATE: 29-JUN-1990

; APPLICATION NUMBER: US 07/457,196

; FILING DATE: 29-DEC-1989

; APPLICATION NUMBER: US 07/390,901

; FILING DATE: 08-AUG-1989

; ATTORNEY/AGENT INFORMATION:

; NAME: Cserr, Luann

; REGISTRATION NUMBER: 31,822

; REFERENCE/DOCKET NUMBER: GI 5190

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: (617) 876-1170

; TELEFAX: (617) 876-5851

; INFORMATION FOR SEQ ID NO: 60:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 1320 amino acids

; TYPE: amino acid

; TOPOLOGY: linear

; MOLECULE TYPE: protein

; SEQUENCE DESCRIPTION: SEQ ID NO: 60:

US-10-124-557-60

Query Match 95.3%; Score 5607.8; DB 13; Length 1320;

Best Local Similarity 96.2%; Pred. No. 1.1e-137;

Matches 1056; Conservative 0; Mismatches 0; Indels 42; Gaps 2;

QY 1 MAWKTLPIYLLLLSVFVIQVSSQDLSSCAGRCGYSRDATA NCYDNCQHYMECCPDF 60

Db 1 MAWKTLPIYLLLLSVFVIQVSSQ----- 25

QY 61 KRVTCTAELSCGRCEPESFERGECDCDAQCKYDKCCPDYESFCAEVHNPTSPSSKKAP 120

Db 26 -----ELSCKGRCEPESFERGECDCDAQCKYDKCCPDYESFCAEVHNPTSPSSKKAP 79

QY 121 PPSGASQTIKSTTKRSPKPPNKKTKKVIIESEITEKVONKKNRTKKKTKPKPPVVD 180

Db 80 PPSGASQTIKSTTKRSPKPPNKKTKKVIIESEITE -VKONKKNRTKKKTKPKPPVVD 138

QY 181 GSGLDNGDFKVTTPDSTTQHNKYSTSPKITTAKPINRPSLPNSDTSKETSITVNKET 240

Db 139 GSGLONGFKVTPDSTSTQHNKSTSPKITTAKINPRPSLPNSDSKSTSTVNKET 198
Qy 241 TVETKETITTTNKQSTGDKETTSKAKETQSIKTSKDLAPTSKVLAKPTPKAETTTKGP 300
Db 199 TVETKETITTTNKQSTGDKETTSKAKETQSIKTSKDLAPTSKVLAKPTPKAETTTKGP 258
Qy 301 ALTTTKEPTPTPKPASTTPKEPTPTTKSAPTTKSAPTTPKPASTTPKSAPTTPKPASTTK 360
Db 259 ALTTTKEPTPTPKPASTTPKEPTPTTKSAPTTKSAPTTPKPASTTPKSAPTTPKPASTTK 318
Qy 361 EPAPTTPKPAPTTTKPAPTTTKSAPTTPKPAPTTPKPAPTTPKPAPTTPKPAPTTPK 420
Db 319 EPAPTTPKPAPTTTKPAPTTTKSAPTTPKPAPTTPKPAPTTPKPAPTTPKPAPTTPK 378
Qy 421 TPKEAPTTKPAPTTPKPAPTTPKPAPTTPKPAPTTPKPAPTTPKPAPTTPKPAPTTPK 480
Db 379 TPKEAPTTKPAPTTPKPAPTTPKPAPTTPKPAPTTPKPAPTTPKPAPTTPKPAPTTPK 438
Qy 481 APPTTKSAPTTTKPAPTTTKSAPTTPKPAPTTPKPAPTTPKPAPTTPKPAPTTPK 540
Db 439 APPTTKSAPTTTKPAPTTTKSAPTTPKPAPTTPKPAPTTPKPAPTTPKPAPTTPK 498
Qy 541 EPAPTTPKPAPTTTKPAPTTPKPAPTTPKPAPTTPKPAPTTPKPAPTTPKPAPTTPK 600
Db 499 EPAPTTPKPAPTTTKPAPTTPKPAPTTPKPAPTTPKPAPTTPKPAPTTPKPAPTTPK 558
Qy 601 TPPEELAPTTPEPTPTTPKEAPTTPKAAAPNTKPAPTTPKPAPTTPKPAPTTPK 660
Db 559 TPPEELAPTTPEPTPTTPKEAPTTPKAAAPNTKPAPTTPKPAPTTPKPAPTTPK 618
Qy 661 ETAPTTPKGAPTTLKPAPTTPKPAKELAPTTKPTSTSTSDKPAPTTPK 720
Db 619 ETAPTTPKGAPTTLKPAPTTPKPAKELAPTTKPTSTSTSDKPAPTTPK 678
Qy 721 KEPAPTTPKEAPTTPKGAPTTLKPAPTTPKPAKELAPTTKPTSTSTSDKPAPTTP 780
Db 679 KEPAPTTPKEAPTTPKGAPTTLKPAPTTPKPAKELAPTTKPTSTSTSDKPAPTTP 738
Qy 781 PKETAPTTKEAPTTPKKPAPTTPETPTPTSEVSTPTTKEPTTIHKSPDESPELSA 840
Db 739 PKETAPTTKEAPTTPKKPAPTTPETPTPTSEVSTPTTKEPTTIHKSPDESPELSA 798
Qy 841 EPTPKALENSKPEGVPTTKTAAKPEMTTAKOKTERDLRTPTTAAKPKMTKTA 900
Db 799 EPTPKALENSKPEGVPTTKTAAKPEMTTAKOKTERDLRTPTTAAKPKMTKTA 858
Qy 901 TTTEKTTESKITATTTQVSTTTQDTPPKITTLKTTTLAPKVTITTKTITTEIMNKE 960
Db 859 TTTEKTTESKITATTTQVSTTTQDTPPKITTLKTTTLAPKVTITTKTITTEIMNKE 918
Qy 961 ETAKPKDRATNSKATTPKPKQKTKAPKPTSTKPKMTPRVRKPTTPTPRKMTSTMPEL 1020
Db 919 ETAKPKDRATNSKATTPKPKQKTKAPKPTSTKPKMTPRVRKPTTPTPRKMTSTMPEL 978
Qy 1021 NPTSRFAEAMLOTTTPRPNOTNSKLVNPKSEDAGAGETPHMLLRHVEMPEVTPDM 1080
Db 979 NPTSRFAEAMLOTTTPRPNOTNSKLVNPKSEDAGAGETPHMLLRHVEMPEVTPDM 1038
Qy 1081 DYLPRVFNQGIINPMLS 1098
Db 1039 DYLPRVFNQGIINPMLS 1056

RESULT 12
US-10-124-557-84
; Sequence 84, Application US/10124557
; Publication No. US20020137894A1
; GENERAL INFORMATION:
; APPLICANT: Turner, Katherine
; Clark, Stephen C.
; Jacobs, Kenneth
; Hewick, Rodney M.

Gesner, Thomas G.
TITLE OF INVENTION: Megakaryocyte Stimulating Factors
NUMBER OF SEQUENCES: 143
CORRESPONDENCE ADDRESS:
ADDRESSEE: Genetics Institute, Inc.
STREET: 87 Cambridgepark Drive
CITY: Cambridge
STATE: Massachusetts
COUNTRY: U.S.A.
ZIP: 02140
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/10/124,557
FILING DATE: 16-Apr-2002
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/643,502
FILING DATE: 18-JAN-1991
APPLICATION NUMBER: US 07/546,114
FILING DATE: 29-JUN-1990
APPLICATION NUMBER: US 07/457,196
FILING DATE: 29-DEC-1989
APPLICATION NUMBER: US 07/390,901
FILING DATE: 08-AUG-1989
ATTORNEY/AGENT INFORMATION:
NAME: Cseri, Luann
REGISTRATION NUMBER: 31,822
REFERENCE/DOCKET NUMBER: GI 5190
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617)876-1170
TELEFAX: (617)876-5851
INFORMATION FOR SEQ ID NO: 84:
SEQUENCE CHARACTERISTICS:
LENGTH: 1022 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
SEQUENCE DESCRIPTION: SEQ ID NO: 84:
US-10-124-557-84

Query Match 92.9%; Score 5466.9; DB 13; Length 1022;
Best Local Similarity 95.1%; Pred. No. 3.7e-134;
Matches 1020; Conservative 0; Mismatches 2; Indels 51; Gaps 2;
Qy 26 DLSSCAGRCGEGYSDATCNCYNCQHYMECCPDFKVCTAEALSCKGRCFESPERGRECD 85
Db 1 DLSSCAGRCGEGYSDATCNCYNCQHYMECCPDFKVC--ELSCKGRCFESPERGRECD 58
Qy 86 CDAQCKYDKCCPDYBSFCAEVHNPTSPSSKKAAPPSSGASQTIKSTKSPKPNKKKT 145
Db 59 CDAQCKYDKCCPDYBSFCAE----- 80
Qy 146 KKVTESEIETKVXDNKNKRTKKKTPKPPVVDAGSLDNGDPKVTTPDTSTTQHNKVS 205
Db 81 -----AVXDNKNKRTKKKTPKPPVVDAGSLDNGDPKVTTPDTSTTQHNKVS 129
Qy 206 TSPKITTAKINPRPSLPNSDTSKETSLLVKNKETTIVETKTTTNKQSTGDKETTS 265
Db 130 TSPKITTAKINPRPSLPNSDTSKETSLLVKNKETTIVETKTTTNKQSTGDKETTS 189
Qy 266 KETOSIEKTSKDLAPTSKVLAKPTPKAETTTKGPALTTKPEPTTPKPASTTPKPT 325
Db 190 KETOSIEKTSKDLAPTSKVLAKPTPKAETTTKGPALTTKPREPTTPKPASTTPKPT 249
Qy 326 PTTIKSAPTTKPAPTTTKSAPTTPKPAPTTPKPAPTTPKPAPTTPKPAPTTTK 385
Db 250 PTTIKSAPTTKPAPTTTKSAPTTPKPAPTTPKPAPTTPKPAPTTPKPAPTTTK 309
Qy 386 APTTKPBPAPTTPKKPAPTTPKPAPTTPKPAPTTPKPAPTTPKPAPTTPKPAPTAP 445

Db 553 TTPBELAPTTPEPTTPPEPATTPKQAAAPNTPKSPAPTTPEPATTPKEPATTPK 612
Qy 661 ETAPTPKGTAPTTKSPAPTTPKKAPKELAPTTTKEPTSTTSQKAPPTPKGTATTP 720
Db 613 ETAPTPKGTAPTTKSPAPTTPKKAPKELAPTTTKEPTSTTSQKAPPTPKGTATTP 672
Qy 721 KEPAPTTPEPATTPKGTATTTKELAPTTPKKAPKELAPTTTKEPTSTTSQKAPPT 780
Db 673 KEPAPTTPEPATTPKGTATTTKELAPTTPKKAPKELAPTTTKEPTSTTSQKAPPT 732
Qy 781 PKETAPTTPEPATTPPKKAPPTPETPPPTTSEVSTPTTTKEPTTHKSPDESTPBLA 840
Db 733 PKETAPTTPEPATTPPKKAPPTPETPPPTTSEVSTPTTTKEPTTHKSPDESTPBLA 792
Qy 841 EPTPKALENSKPEKCVPTTKPAATKPEMTTAKDKTTERDLRTPTETTHAAPKWTETA 900
Db 793 EPTPKALENSKPKBPVPTTKPAATKPEMTTAKDKTTERDLRTPTETTHAAPKWTETA 852
Qy 901 TTTETKTESKITATTTQVTSSTTTQDTTPFKITTTLLKTTTLAPKVTITTKTITTTIMNKPE 960
Db 853 TTTETKTESKITATTTQVTSSTTTQDTTPFKITTTLLKTTTLAPKVTITTKTITTTIMNKPE 912
Qy 961 ETAKPKDRATNSKATTPKQKPTKAPKXPTSTKKPKTMPRVKPKTTPTRKMTSTMPBL 1020
Db 913 ETAKPKDRATNSKATTPKQKPTKAPKXPTSTKKPKTMPRVKPKTTPTRKMTSTMPBL 972
Qy 1021 NPTSRIAEAMLOTTTRPNQTNKSLVEVNPKESEDAGGAEGETPHMLLRPHVFMPEVTPDM 1080
Db 973 NPTSRIAEAMLOTTTRPNQTNKSLVEVNPKESEDAGGAEGETPHMLLRPHVFMPEVTPDM 1032
Qy 1081 DYLPRVNPQGIINPMLS 1098
Db 1033 DYLPRVNPQGIINPMLS 1050

RESULT 14
US-10-124-557-58
; Sequence 58, Application US/10124557
; Publication No. US20020137894A1
; GENERAL INFORMATION:
; APPLICANT: Turner, Katherine
; Jacobs, Kenneth
; Hewick, Rodney M.
; Gesner, Thomas G.
; TITLE OF INVENTION: Megakaryocyte Stimulating Factors
; NUMBER OF SEQUENCES: 143
; CORRESPONDENCE ADDRESS:
; ADDRESS: Genetics Institute, Inc.
; STREET: 87 Cambridgepark Drive
; CITY: Cambridge
; STATE: Massachusetts
; COUNTRY: U.S.A.
; Zip: 02140
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/10/124,557
; FILING DATE: 16-APR-2002
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/643,502
; FILING DATE: 18-JAN-1991
; APPLICATION NUMBER: US 07/546,114
; FILING DATE: 29-JUN-1990
; APPLICATION NUMBER: US 07/457,196
; FILING DATE: 29-DEC-1989
; APPLICATION NUMBER: US 07/390,901
; FILING DATE: 08-AUG-1989

ATTORNEY/AGENT INFORMATION:
NAME: Cseri, Luann
REGISTRATION NUMBER: 31,822
REFERENCE/DOCKET NUMBER: GI 5190
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617)876-1170
TELEFAX: (617)876-5851
INFORMATION FOR SEQ ID NO: 58:
SEQUENCE CHARACTERISTICS:
LENGTH: 1049 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
SEQUENCE DESCRIPTION: SEQ ID NO: 58:
US-10-124-557-58
Query Match 91.4%; Score 5383.1; DB 13; Length 1049;
Best Local Similarity 92.5%; Pred. No. 5.7e-132;
Matches 1046; Conservative 7; Mismatches 26; Indels 49; Gaps 2;
Qy 1 MAWKTLPIYLLLLSVFVIQQVSSQDLSSCAGRCGEGYSRDATCNDYNCQHYMECCPDF 60
Db 1 MAWKTLPIYLLLLSVFVIQQVSSQ----- 25
Qy 61 KRYCTAELSCGRCFSEFERGECDDAOCKYDKCCPDYEGFCAEVHNPTSPSSSKAP 120
Db 26 -----ELSCGRCFSEFERGECDDAOCKYDKCCPDYEGFCAEHVSSENQESSSS 79
Qy 121 PPSGASQTKITTKRSPKPPNKKTKVIESBEITEKVNDKNKRTKKKTPKPPVDEA 180
Db 80 SSSSSSTIWKIKSSKNANRELQKLL-----KVKDNKNKRTKKKTPKPPVDEA 131
Qy 181 GSGLDNGDFKVTTPDTSTTHNKNVSTSPKLTITAKPINRPSLPNSDTSKETSITVKNK 240
Db 132 GSGLDNGDFKVTTPDTSTTHNKNVSTSPKLTITAKPINRPSLPNSDTSKETSITVKNK 191
Qy 241 TVETKETTITNKOTSTDGKEKTTSAKETOSIEKTSADLAPTSSKVLAKTPPKAETTKGP 300
Db 192 TVETKETTITNKOTSTDGKEKTTSAKETOSIEKTSADLAPTSSKVLAKTPPKAETTKGP 251
Qy 301 ALATPKPEPTTPKPEASTTPKPEPTTTPKSAFTTPKPEAPTTKSAFTTPKPEAPTTTK 360
Db 252 ALATPKPEPTTPKPEASTTPKPEPTTTPKSAFTTPKPEAPTTKSAFTTPKPEAPTTTK 311
Qy 361 EPAPTTPEPATTTPEPATTTKSAFTTPKPEAPTTKSAFTTPKPEAPTTKPEPTTP 420
Db 312 EPAPTTPEPATTTPEPATTTKSAFTTPKPEAPTTKSAFTTPKPEAPTTKPEPTTP 371
Qy 421 TPKEPATTTPEPATTTPEPATTTKSAFTTPKPEAPTTKSAFTTPKPEAPTTKPEPTTP 480
Db 372 TPKEPATTTPEPATTTPEPATTTKSAFTTPKPEAPTTKSAFTTPKPEAPTTKPEPTTP 431
Qy 481 APPTTKSAFTTPKPEAPTTKSAFTTPKPEAPTTKSAFTTPKPEAPTTKPEAPTTK 540
Db 432 APPTTKSAFTTPKPEAPTTKSAFTTPKPEAPTTKSAFTTPKPEAPTTKPEAPTTK 491
Qy 541 EPAPTTPEPATTTKSAFTTPKPEAPTTKSAFTTPKSAFTTPKSAFTTPKSAFTTP 600
Db 492 EPAPTTPEPATTTKSAFTTPKPEAPTTKSAFTTPKSAFTTPKSAFTTPKSAFTTP 551
Qy 601 TTPBELAPTTPEPTTPPEPATTTKSAFTTPKSAFTTPKSAFTTPKSAFTTPKSAFTTP 660
Db 552 TTPBELAPTTPEPTTPPEPATTTKSAFTTPKSAFTTPKSAFTTPKSAFTTPKSAFTTP 611
Qy 661 ETAPTTKGTAPTTKELAPTTPKKAPKELAPTTTKEPTSTTSQKAPPTPKGTATTP 720
Db 612 ETAPTTKGTAPTTKELAPTTPKKAPKELAPTTTKEPTSTTSQKAPPTPKGTATTP 671
Qy 721 KEPAPTTPEPATTPKGTATTTKELAPTTPKKAPKELAPTTTKEPTSTTSQKAPPT 780
Db 672 KEPAPTTPEPATTPKGTATTTKELAPTTPKKAPKELAPTTTKEPTSTTSQKAPPT 731
Qy 781 PKETAPTTPEPATTTKGTATTTKELAPTTPKKAPKELAPTTTKEPTSTTSQKAPPT 840

| | | | |
|----|------|---|------|
| Db | 912 | ETAKPKDRATNSKATTEPKOKTKAPKKPTSTKKPKTTPRVRKPKTTPRKOVTSMPEL | 971 |
| Qy | 1021 | NPTSRIAEAMLOTTTRENQTPNSKLVENVNPKSEDAAGAGETPHMLLRPHVWPEVTPDM | 1080 |
| Db | 972 | NPTSRIAEAMLOTTTRENQTPNSKLVENVNPKSEDAAGAGETPHMLLRPHVWPEVTPDM | 1031 |
| Qy | 1081 | DYLPRVENQGIILNPMLS | 1098 |
| Db | 1032 | DYLPRVENQGIILNPMLS | 1049 |

Search completed: October 13, 2004, 11:52:51
Job time : 117.709 secs

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: October 13, 2004, 11:23:49 ; Search time 120.725 Seconds
(without alignments)
5233.063 Million cell updates/sec

Title: SE01-C
Perfect score: 5837
Sequence: 1 MAWKTLPIYLLLLLSVFVIQ.....DMDYLPRVNCGIINPMLS 1098

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.1

Searched: 1825181 seqs, 575374646 residues

Total number of hits satisfying chosen parameters: 1825181

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%
Listing first 45 summaries

Database : UniProt_02: *
1: uniprot_sprot: *
2: uniprot_trembl: *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | DB ID | Description |
|------------|--------|-------------|--------|-------|--------------------|
| 1 | 5872.8 | 99.8 | 1404 | 2 | Q92954 homo sapien |
| 2 | 5863.8 | 99.6 | 1404 | 2 | Q9BX49 |
| 3 | 3204.7 | 54.4 | 933 | 2 | Q6ZM25 |
| 4 | 3204.7 | 54.4 | 933 | 2 | BAD18580 |
| 5 | 2731.5 | 46.4 | 1054 | 2 | Q9JW99 |
| 6 | 1336.9 | 23.7 | 5173 | 1 | MUC2 HUMAN |
| 7 | 1327.5 | 22.5 | 1761 | 2 | Q9VR49 |
| 8 | 1327.5 | 22.5 | 1761 | 2 | Q7KTF6 |
| 9 | 1266.1 | 21.5 | 3150 | 2 | AAS4673 |
| 10 | 1266.1 | 21.5 | 3150 | 2 | Q6SSE6 |
| 11 | 1205.1 | 20.5 | 3409 | 2 | AAS07044 |
| 12 | 1205.1 | 20.5 | 3409 | 2 | SLP1_CLOTM |
| 13 | 1132 | 19.2 | 1684 | 2 | Q6SSE8 |
| 14 | 1124.7 | 19.1 | 3889 | 2 | AAS07042 |
| 15 | 1124.7 | 19.1 | 3889 | 2 | Q8WQ4 |
| 16 | 1115.3 | 18.9 | 1349 | 2 | Q8WRQ4 |
| 17 | 1108.1 | 18.8 | 3432 | 2 | Q8IR51 |
| 18 | 1108.1 | 18.8 | 3432 | 2 | Q8IR52 |
| 19 | 1100.7 | 18.7 | 1795 | 2 | Q76894 |
| 20 | 1065.6 | 18.1 | 1079 | 2 | Q9N457 |
| 21 | 1054.7 | 17.9 | 23015 | 2 | Q8IQ18 |
| 22 | 1054.7 | 17.9 | 23015 | 2 | AAAI0358 |
| 23 | 1049.8 | 17.8 | 9234 | 2 | Q7KTP5 |
| 24 | 1049.8 | 17.8 | 9234 | 2 | AAAI0531 |
| 25 | 1047.2 | 17.8 | 2284 | 2 | Q9VPG1 |
| 26 | 1042.9 | 17.7 | 1607 | 2 | Q8H6Q5 |
| 27 | 1042.9 | 17.7 | 1607 | 2 | AP474661 |
| 28 | 1032.6 | 17.5 | 5703 | 1 | MUSB_HUMAN |
| 29 | 1028.4 | 17.5 | 972 | 2 | Q7QKK7 |
| 30 | 1020.9 | 17.3 | 1489 | 2 | Q96449 |
| 31 | 1016.3 | 17.3 | 34350 | 2 | Q8WZ42 |

Q10465 homo sapien
Q20007 caenorhabdi
Q815f5 caenorhabdi
Q815f6 caenorhabdi
Q815f7 caenorhabdi
Q911e8 arabidopsis
Q90670 mus musculus
Q757N5 ashbya goss
AAS2662 ashbya go
Q9vel9 drosophila
Q72884 candida alb
Q9W223 drosophila
Q76718 sus scrofa

ALIGNMENTS

RESULT 1
Q92954 PRELIMINARY; PRT; 1404 AA.
AC Q92954;
DT 01-FEB-1997 (Tremblrel_02, Created)
DT 01-FEB-1997 (Tremblrel_02, Last sequence update)
DT 01-MAR-2004 (Tremblrel_25, Last annotation update)
DE Megakaryocyte stimulating factor.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCSI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Turner K.J., Fitz L.J., Temple P., Jacobs K., Larson D., Leary A.C.,
RA Kelleher K., Giannotti J., Calvetti J., Fitzgerald M., Kriz M.J.,
RA Ferenz C., Grobholz J., Fraser H., Bean K., Norton C.R., Gesner T.,
RA Bhatia S., Kriz R., Hewick R., Clark S.C.;
RT "Purification, Biochemical Characterization, and Cloning of a Novel
RT Megakaryocyte Stimulating Factor that has Megakaryocyte Colony
RT Stimulating Activity.";
RL Blood 78:279-279(1991).
RN [2]
RP SEQUENCE FROM N.A.
RA Merberg D.M., Fitz L.J., Temple P., Giannotti J., Murtha P.,
RA Fitzgerald M., Scaltreto J., Kelleher K., Preissner K., Kriz R.,
RA Jacobs K., Turner K.;
RT "A Comparison of Vitronectin and Megakaryocyte Stimulating Factor.";
RL (In) Preissner K.T., Rosenblatt S., Kost C., Wegerhoff J.,
RL Mosher D.F. (eds.);
RL BIOLOGY OF VITRONECTINS AND THEIR RECEPTORS., pp.45-52, Elsevier
RL Science Publishers B.V. (1993).
RN [3]
RP SEQUENCE FROM N.A.
RA Turner K.J., Fitz L.J., Temple P., Jacobs K., Larson D., Leary A.C.,
RA Kelleher K., Giannotti J., Calvetti J., Fitzgerald M., Kriz M.J.,
RA Ferenz C., Grobholz J., Fraser H., Bean K., Norton C.R., Gesner T.,
RA Bhatia S., Kriz R., Hewick R., Clark S.C.;
RL Submitted (SEP-1996) to the EMBL/GenBank/DBJ databases.
DR EMBL; U70136; AA09089.1; -.
DR HSP; P04004; I0C0.
DR Genew; HGNC:9364; PRG4.
DR GO; GO:0008283; P:cell proliferation; TAS.
DR InterPro; IPR001212; Somatomedin_B.
DR InterPro; IPR001212; Somatomedin_B.
DR Pfam; PF00045; Hemopexin; 2.
DR Pfam; PF00045; Hemopexin; 2.
DR PRINTS; PR00022; Somatomedin_B; 2.
DR PRINTS; PR00022; Somatomedin_B; 2.
DR SMART; SM00201; SO; 2.
DR SMART; SM00201; SO; 2.
DR PROSITE; PS00024; HEMOPEXIN; UNKNOWN_1.
DR PROSITE; PS00524; SOMATOMEDIN_B; 2.
DR PROSITE; PS00524; SOMATOMEDIN_B; 2.
SQ SEQUENCE 1404 AA; 151090 MW; AABD7AD19B35F4F6 CRC64;

```

Query Match      99.8%; Score 5872.8; DB 2; Length 1404;
Best Local Similarity 96.3%; Pred. No. 1.6e-104;
Matches 1096; Conservative 0; Mismatches 0; Indels 42; Gaps 1;

QY 1 MAWKTLPIYLLLSVFIQVSSQDLSSCAGRCGEGYSDATCNCYNCQHYMECCPDF 60
DB 1 MAWKTLPIYLLLSVFIQVSSQDLSSCAGRCGEGYSDATCNCYNCQHYMECCPDF 60
QY 61 KRVCYAEISCKGRCFESFERGECDCDAOCKYDKCCPDYEFCAEVNPTSPSSKKAP 120
DB 61 KRVCYAEISCKGRCFESFERGECDCDAOCKYDKCCPDYEFCAEVNPTSPSSKKAP 120
QY 121 PPSGASQTIKSTTKRSPKPPNKKTKKVIIESEITE----- 156
DB 121 PPSGASQTIKSTTKRSPKPPNKKTKKVIIESEITE----- 156
QY 121 PPSGASQTIKSTTKRSPKPPNKKTKKVIIESEITE----- 156
DB 121 PPSGASQTIKSTTKRSPKPPNKKTKKVIIESEITE----- 156
QY 157 -----KVKDNKNRTKKKTPKPPVVDGAGSLDNGDFKVTTPDST 198
DB 157 -----KVKDNKNRTKKKTPKPPVVDGAGSLDNGDFKVTTPDST 198
QY 181 KIKSSNSAANRELQKLVKDNKNRTKKKTPKPPVVDGAGSLDNGDFKVTTPDST 240
DB 181 KIKSSNSAANRELQKLVKDNKNRTKKKTPKPPVVDGAGSLDNGDFKVTTPDST 240
QY 199 TQHNKVSTSPKITTAKPINRPSLPNSDTSKETSITVKNKETTVEKTTTNTKQSTDG 258
DB 199 TQHNKVSTSPKITTAKPINRPSLPNSDTSKETSITVKNKETTVEKTTTNTKQSTDG 258
QY 241 TQHNKVSTSPKITTAKPINRPSLPNSDTSKETSITVKNKETTVEKTTTNTKQSTDG 300
DB 241 TQHNKVSTSPKITTAKPINRPSLPNSDTSKETSITVKNKETTVEKTTTNTKQSTDG 300
QY 259 KEKTTSAKETQSIIEKTSKADLAPTSKVLAETTKGPAETTKGPAETTKGPAETTKGPA 318
DB 259 KEKTTSAKETQSIIEKTSKADLAPTSKVLAETTKGPAETTKGPAETTKGPAETTKGPA 318
QY 301 KEKTTSAKETQSIIEKTSKADLAPTSKVLAETTKGPAETTKGPAETTKGPAETTKGPA 360
DB 301 KEKTTSAKETQSIIEKTSKADLAPTSKVLAETTKGPAETTKGPAETTKGPAETTKGPA 360
QY 319 TTPKEPTPTTIKSAPTTKGPAETTKGPAETTKGPAETTKGPAETTKGPAETTKGPA 378
DB 319 TTPKEPTPTTIKSAPTTKGPAETTKGPAETTKGPAETTKGPAETTKGPAETTKGPA 378
QY 361 TTPKEPTPTTIKSAPTTKGPAETTKGPAETTKGPAETTKGPAETTKGPAETTKGPA 420
DB 361 TTPKEPTPTTIKSAPTTKGPAETTKGPAETTKGPAETTKGPAETTKGPAETTKGPA 420
QY 379 APTTTKSAPTTKGPAETTKGPAETTKGPAETTKGPAETTKGPAETTKGPAETTKGPA 438
DB 379 APTTTKSAPTTKGPAETTKGPAETTKGPAETTKGPAETTKGPAETTKGPAETTKGPA 438
QY 421 APTTTKSAPTTKGPAETTKGPAETTKGPAETTKGPAETTKGPAETTKGPAETTKGPA 480
DB 421 APTTTKSAPTTKGPAETTKGPAETTKGPAETTKGPAETTKGPAETTKGPAETTKGPA 480
QY 439 BPAPTAPKPAETTKGPAETTKGPAETTKGPAETTKGPAETTKGPAETTKGPAETTKGPA 498
DB 439 BPAPTAPKPAETTKGPAETTKGPAETTKGPAETTKGPAETTKGPAETTKGPAETTKGPA 498
QY 481 BPAPTAPKPAETTKGPAETTKGPAETTKGPAETTKGPAETTKGPAETTKGPAETTKGPA 540
DB 481 BPAPTAPKPAETTKGPAETTKGPAETTKGPAETTKGPAETTKGPAETTKGPAETTKGPA 540
QY 499 TTKSAPTTKGPAETTKGPAETTKGPAETTKGPAETTKGPAETTKGPAETTKGPAETTKGPA 558
DB 499 TTKSAPTTKGPAETTKGPAETTKGPAETTKGPAETTKGPAETTKGPAETTKGPAETTKGPA 558
QY 541 TTKSAPTTKGPAETTKGPAETTKGPAETTKGPAETTKGPAETTKGPAETTKGPAETTKGPA 600
DB 541 TTKSAPTTKGPAETTKGPAETTKGPAETTKGPAETTKGPAETTKGPAETTKGPAETTKGPA 600
QY 559 APTAPKPAETTKGPAETTKGPAETTKGPAETTKGPAETTKGPAETTKGPAETTKGPAETTKGPA 618
DB 559 APTAPKPAETTKGPAETTKGPAETTKGPAETTKGPAETTKGPAETTKGPAETTKGPAETTKGPA 618
QY 601 APTAPKPAETTKGPAETTKGPAETTKGPAETTKGPAETTKGPAETTKGPAETTKGPAETTKGPA 660
DB 601 APTAPKPAETTKGPAETTKGPAETTKGPAETTKGPAETTKGPAETTKGPAETTKGPAETTKGPA 660
QY 619 PEEAPTTPKAAANPTKPAETTKGPAETTKGPAETTKGPAETTKGPAETTKGPAETTKGPA 678
DB 619 PEEAPTTPKAAANPTKPAETTKGPAETTKGPAETTKGPAETTKGPAETTKGPAETTKGPA 678
QY 661 PEEAPTTPKAAANPTKPAETTKGPAETTKGPAETTKGPAETTKGPAETTKGPAETTKGPA 720
DB 661 PEEAPTTPKAAANPTKPAETTKGPAETTKGPAETTKGPAETTKGPAETTKGPAETTKGPA 720
QY 679 APTTPKPAKPAETTKGPAETTKGPAETTKGPAETTKGPAETTKGPAETTKGPAETTKGPA 738
DB 679 APTTPKPAKPAETTKGPAETTKGPAETTKGPAETTKGPAETTKGPAETTKGPAETTKGPA 738
QY 721 APTTPKPAKPAETTKGPAETTKGPAETTKGPAETTKGPAETTKGPAETTKGPAETTKGPA 780
DB 721 APTTPKPAKPAETTKGPAETTKGPAETTKGPAETTKGPAETTKGPAETTKGPAETTKGPA 780
QY 739 TAPTTLKPAETTKGPAETTKGPAETTKGPAETTKGPAETTKGPAETTKGPAETTKGPAETTKGPA 798
DB 739 TAPTTLKPAETTKGPAETTKGPAETTKGPAETTKGPAETTKGPAETTKGPAETTKGPAETTKGPA 798
QY 781 TAPTTLKPAETTKGPAETTKGPAETTKGPAETTKGPAETTKGPAETTKGPAETTKGPAETTKGPA 840
DB 781 TAPTTLKPAETTKGPAETTKGPAETTKGPAETTKGPAETTKGPAETTKGPAETTKGPAETTKGPA 840
QY 799 KPAETTPETPPPTTSVSTPTTKGPAETTKGPAETTKGPAETTKGPAETTKGPAETTKGPA 858
DB 799 KPAETTPETPPPTTSVSTPTTKGPAETTKGPAETTKGPAETTKGPAETTKGPAETTKGPA 858
QY 841 KPAETTPETPPPTTSVSTPTTKGPAETTKGPAETTKGPAETTKGPAETTKGPAETTKGPA 900
DB 841 KPAETTPETPPPTTSVSTPTTKGPAETTKGPAETTKGPAETTKGPAETTKGPAETTKGPA 900
QY 859 TKTAAAKPEMTTAKDKTTERDLRTTPTTAAKMTKETATTTTETKTTESKLTATTTQV 918
DB 859 TKTAAAKPEMTTAKDKTTERDLRTTPTTAAKMTKETATTTTETKTTESKLTATTTQV 918
QY 901 TKTAAAKPEMTTAKDKTTERDLRTTPTTAAKMTKETATTTTETKTTESKLTATTTQV 960
DB 901 TKTAAAKPEMTTAKDKTTERDLRTTPTTAAKMTKETATTTTETKTTESKLTATTTQV 960
QY 919 TSTTTQDTPFKITLTKTTLAPKVTITTKTITTTTEIMNKPEETAKPKDRATNSKATTPK 978
DB 919 TSTTTQDTPFKITLTKTTLAPKVTITTKTITTTTEIMNKPEETAKPKDRATNSKATTPK 978
QY 961 TSTTTQDTPFKITLTKTTLAPKVTITTKTITTTTEIMNKPEETAKPKDRATNSKATTPK 1020
DB 961 TSTTTQDTPFKITLTKTTLAPKVTITTKTITTTTEIMNKPEETAKPKDRATNSKATTPK 1020
QY 979 PQKPTAKPKPTSTKPKTKMTPRVKPKTTTTPKMTSTMPELNPTSPRIEAMLOTTTREN 1038
DB 979 PQKPTAKPKPTSTKPKTKMTPRVKPKTTTTPKMTSTMPELNPTSPRIEAMLOTTTREN 1038

```

```

Db 1021 PQKPTAKPKPTSTKPKTKMTPRVKPKTTTTPKMTSTMPELNPTSPRIEAMLOTTTREN 1080
QY 1039 QTPNSKLVEYNPKSEADAGAGETPHMLRPHVFMPEVTPDMDYLPRVNOGIIINPMLS 1098
DB 1081 QTPNSKLVEYNPKSEADAGAGETPHMLRPHVFMPEVTPDMDYLPRVNOGIIINPMLS 1140

RESULT 2
Q9BX49 PRELIMINARY; PRT; 1404 AA.
ID Q9BX49 AC
Q9BX49;
DT 01-JUN-2001 (T-EMBLrel. 17, Created)
DT 01-JUN-2001 (T-EMBLrel. 17, Last sequence update)
DT 01-MAR-2004 (T-EMBLrel. 26, Last annotation update)
DE BG174L6.2 (MSF: megakaryocyte stimulating factor ).
GN Name=BG174L6.2;
OS Homo sapiens (Human);
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Wray P.;
RL Submitted (JUL-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AL133553; CAC36090.1; -.
DR HSP; P04004; LOC0.
DR InterPro; IPR000595; Hemopexin.
DR InterPro; IPR001212; Somatomedin_B.
DR Pfam; PF00045; Hemopexin; 2.
DR Pfam; PF01033; Somatomedin_B; 2.
DR PRINTS; PR00022; SOMATOMEDINB.
DR SMART; SM00120; HX; 2.
DR SMART; SMC0201; SO; 2.
DR PROSITE; PS00024; HEMOPEXIN; UNKNOWN_1.
DR PROSITE; PS00024; SOMATOMEDIN_B; 2.
SQ SEQUENCE 1404 AA; 151076 MW; 782A11746B3FDBE5 CRC64;

Query Match      99.6%; Score 5863.8; DB 2; Length 1404;
Best Local Similarity 96.1%; Pred. No. 2.4e-104;
Matches 1096; Conservative 0; Mismatches 2; Indels 42; Gaps 1;

QY 1 MAWKTLPIYLLLSVFIQVSSQDLSSCAGRCGEGYSDATCNCYNCQHYMECCPDF 60
DB 1 MAWKTLPIYLLLSVFIQVSSQDLSSCAGRCGEGYSDATCNCYNCQHYMECCPDF 60
QY 61 KRVCYAEISCKGRCFESFERGECDCDAOCKYDKCCPDYEFCAEVNPTSPSSKKAP 120
DB 61 KRVCYAEISCKGRCFESFERGECDCDAOCKYDKCCPDYEFCAEVNPTSPSSKKAP 120
QY 121 PPSGASQTIKSTTKRSPKPPNKKTKKVIIESEITE----- 156
DB 121 PPSGASQTIKSTTKRSPKPPNKKTKKVIIESEITE----- 156
QY 121 PPSGASQTIKSTTKRSPKPPNKKTKKVIIESEITE----- 156
DB 121 PPSGASQTIKSTTKRSPKPPNKKTKKVIIESEITE----- 156
QY 157 -----KVKDNKNRTKKKTPKPPVVDGAGSLDNGDFKVTTPDST 198
DB 157 -----KVKDNKNRTKKKTPKPPVVDGAGSLDNGDFKVTTPDST 198
QY 181 KIKSSNSAANRELQKLVKDNKNRTKKKTPKPPVVDGAGSLDNGDFKVTTPDST 240
DB 181 KIKSSNSAANRELQKLVKDNKNRTKKKTPKPPVVDGAGSLDNGDFKVTTPDST 240
QY 199 TQHNKVSTSPKITTAKPINRPSLPNSDTSKETSITVKNKETTVEKTTTNTKQSTDG 258
DB 199 TQHNKVSTSPKITTAKPINRPSLPNSDTSKETSITVKNKETTVEKTTTNTKQSTDG 258
QY 241 TQHNKVSTSPKITTAKPINRPSLPNSDTSKETSITVKNKETTVEKTTTNTKQSTDG 300
DB 241 TQHNKVSTSPKITTAKPINRPSLPNSDTSKETSITVKNKETTVEKTTTNTKQSTDG 300
QY 259 KEKTTSAKETQSIIEKTSKADLAPTSKVLAETTKGPAETTKGPAETTKGPAETTKGPA 318
DB 259 KEKTTSAKETQSIIEKTSKADLAPTSKVLAETTKGPAETTKGPAETTKGPAETTKGPA 318
QY 301 KEKTTSAKETQSIIEKTSKADLAPTSKVLAETTKGPAETTKGPAETTKGPAETTKGPA 360
DB 301 KEKTTSAKETQSIIEKTSKADLAPTSKVLAETTKGPAETTKGPAETTKGPAETTKGPA 360
QY 319 TTPKEPTPTTIKSAPTTKGPAETTKGPAETTKGPAETTKGPAETTKGPAETTKGPA 378
DB 319 TTPKEPTPTTIKSAPTTKGPAETTKGPAETTKGPAETTKGPAETTKGPAETTKGPA 378
QY 361 TTPKEPTPTTIKSAPTTKGPAETTKGPAETTKGPAETTKGPAETTKGPAETTKGPA 420
DB 361 TTPKEPTPTTIKSAPTTKGPAETTKGPAETTKGPAETTKGPAETTKGPAETTKGPA 420
QY 379 APTTTKSAPTTKGPAETTKGPAETTKGPAETTKGPAETTKGPAETTKGPAETTKGPA 438
DB 379 APTTTKSAPTTKGPAETTKGPAETTKGPAETTKGPAETTKGPAETTKGPAETTKGPA 438
QY 421 APTTTKSAPTTKGPAETTKGPAETTKGPAETTKGPAETTKGPAETTKGPAETTKGPA 480
DB 421 APTTTKSAPTTKGPAETTKGPAETTKGPAETTKGPAETTKGPAETTKGPAETTKGPA 480
QY 439 BPAPTAPKPAETTKGPAETTKGPAETTKGPAETTKGPAETTKGPAETTKGPAETTKGPA 498
DB 439 BPAPTAPKPAETTKGPAETTKGPAETTKGPAETTKGPAETTKGPAETTKGPAETTKGPA 498
QY 481 BPAPTAPKPAETTKGPAETTKGPAETTKGPAETTKGPAETTKGPAETTKGPAETTKGPA 540
DB 481 BPAPTAPKPAETTKGPAETTKGPAETTKGPAETTKGPAETTKGPAETTKGPAETTKGPA 540
QY 499 TTKSAPTTKGPAETTKGPAETTKGPAETTKGPAETTKGPAETTKGPAETTKGPAETTKGPA 558
DB 499 TTKSAPTTKGPAETTKGPAETTKGPAETTKGPAETTKGPAETTKGPAETTKGPAETTKGPA 558
QY 541 TTKSAPTTKGPAETTKGPAETTKGPAETTKGPAETTKGPAETTKGPAETTKGPAETTKGPA 600
DB 541 TTKSAPTTKGPAETTKGPAETTKGPAETTKGPAETTKGPAETTKGPAETTKGPAETTKGPA 600
QY 559 APTAPKPAETTKGPAETTKGPAETTKGPAETTKGPAETTKGPAETTKGPAETTKGPAETTKGPA 618
DB 559 APTAPKPAETTKGPAETTKGPAETTKGPAETTKGPAETTKGPAETTKGPAETTKGPAETTKGPA 618
QY 601 APTAPKPAETTKGPAETTKGPAETTKGPAETTKGPAETTKGPAETTKGPAETTKGPAETTKGPA 660
DB 601 APTAPKPAETTKGPAETTKGPAETTKGPAETTKGPAETTKGPAETTKGPAETTKGPAETTKGPA 660
QY 619 PEEAPTTPKAAANPTKPAETTKGPAETTKGPAETTKGPAETTKGPAETTKGPAETTKGPA 678
DB 619 PEEAPTTPKAAANPTKPAETTKGPAETTKGPAETTKGPAETTKGPAETTKGPAETTKGPA 678
QY 661 PEEAPTTPKAAANPTKPAETTKGPAETTKGPAETTKGPAETTKGPAETTKGPAETTKGPA 720
DB 661 PEEAPTTPKAAANPTKPAETTKGPAETTKGPAETTKGPAETTKGPAETTKGPAETTKGPA 720
QY 679 APTTPKPAKPAETTKGPAETTKGPAETTKGPAETTKGPAETTKGPAETTKGPAETTKGPA 738
DB 679 APTTPKPAKPAETTKGPAETTKGPAETTKGPAETTKGPAETTKGPAETTKGPAETTKGPA 738
QY 721 APTTPKPAKPAETTKGPAETTKGPAETTKGPAETTKGPAETTKGPAETTKGPAETTKGPA 780
DB 721 APTTPKPAKPAETTKGPAETTKGPAETTKGPAETTKGPAETTKGPAETTKGPAETTKGPA 780
QY 739 TAPTTLKPAETTKGPAETTKGPAETTKGPAETTKGPAETTKGPAETTKGPAETTKGPAETTKGPA 798
DB 739 TAPTTLKPAETTKGPAETTKGPAETTKGPAETTKGPAETTKGPAETTKGPAETTKGPAETTKGPA 798
QY 781 TAPTTLKPAETTKGPAETTKGPAETTKGPAETTKGPAETTKGPAETTKGPAETTKGPAETTKGPA 840
DB 781 TAPTTLKPAETTKGPAETTKGPAETTKGPAETTKGPAETTKGPAETTKGPAETTKGPAETTKGPA 840
QY 799 KPAETTPETPPPTTSVSTPTTKGPAETTKGPAETTKGPAETTKGPAETTKGPAETTKGPA 858
DB 799 KPAETTPETPPPTTSVSTPTTKGPAETTKGPAETTKGPAETTKGPAETTKGPAETTKGPA 858
QY 841 KPAETTPETPPPTTSVSTPTTKGPAETTKGPAETTKGPAETTKGPAETTKGPAETTKGPA 900
DB 841 KPAETTPETPPPTTSVSTPTTKGPAETTKGPAETTKGPAETTKGPAETTKGPAETTKGPA 900
QY 859 TKTAAAKPEMTTAKDKTTERDLRTTPTTAAKMTKETATTTTETKTTESKLTATTTQV 918
DB 859 TKTAAAKPEMTTAKDKTTERDLRTTPTTAAKMTKETATTTTETKTTESKLTATTTQV 918
QY 901 TKTAAAKPEMTTAKDKTTERDLRTTPTTAAKMTKETATTTTETKTTESKLTATTTQV 960
DB 901 TKTAAAKPEMTTAKDKTTERDLRTTPTTAAKMTKETATTTTETKTTESKLTATTTQV 960
QY 919 TSTTTQDTPFKITLTKTTLAPKVTITTKTITTTTEIMNKPEETAKPKDRATNSKATTPK 978
DB 919 TSTTTQDTPFKITLTKTTLAPKVTITTKTITTTTEIMNKPEETAKPKDRATNSKATTPK 978
QY 961 TSTTTQDTPFKITLTKTTLAPKVTITTKTITTTTEIMNKPEETAKPKDRATNSKATTPK 1020
DB 961 TSTTTQDTPFKITLTKTTLAPKVTITTKTITTTTEIMNKPEETAKPKDRATNSKATTPK 1020
QY 979 PQKPTAKPKPTSTKPKTKMTPRVKPKTTTTPKMTSTMPELNPTSPRIEAMLOTTTREN 1038
DB 979 PQKPTAKPKPTSTKPKTKMTPRVKPKTTTTPKMTSTMPELNPTSPRIEAMLOTTTREN 1038

```



```

Db 430 TKTFAATKEMTTAKDKTTERDLRTPETTTAAAPKMTKETATTTKTESKITATTQV 489
QY 919 TSTTTQDTPFKITLTKTTLAPKVTITTKKTIITTEIMNKPEETAKPKDRATNSKATTPK 978
Db 490 TSTTTQDTPFKITLTKTTLAPKVTITTKKTIITTEIMNKPEETAKPKDRATNSKATTPK 549
QY 979 POKPTKAPKPKPTSTKPKTMPRVKPKTTPTRKMTSTMPELNPTSRIAEAMLQTTTRPN 1038
Db 550 POKPTKAPKPKPTSTKPKTMPRVKPKTTPTRKMTSTMPELNPTSRIAEAMLQTTTRPN 609
QY 1039 QTPNSKLVNPKSEDAGAGETPHMLLRPHVFMPEVTPDMDYLPVRVNOGIIINPMLS 1098
Db 610 QTPNSKLVNPKSEDAGAGETPHMLLRPHVFMPEVTPDMDYLPVRVNOGIIINPMLS 669

RESULT 4
ID BADI8580 PRELIMINARY; PRT; 933 AA.
AC BADI8580;
DT 12-MAY-2004 (TREMELrel. 27, Created)
DT 12-MAY-2004 (TREMELrel. 27, Last sequence update)
DE CDNA FLJ16561 fis, clone SYN04003981, moderately similar to Homo
DE sapiens proteoglycan 4, (megakaryocyte stimulating factor, articular
DE pericarditis syndrome) (PRG4).
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Synovial membrane tissue;
RA Ota T., Nakagawa S., Senoh A., Mizuguchi H., Inagaki H., Sugiyama T.,
RA Irie R., Otsuki T., Sato H., Wakamatsu A., Ishii S., Yamamoto J.,
RA Isono Y., Kawai-Hio Y., Saito K., Nishikawa T., Kimura K.,
RA Yamashita H., Matsuo K., Nakamura Y., Sekine M., Kikuchi H., Kanda K.,
RA Wagatsuma M., Murakawa K., Kanehori K., Takahashi-Fujii A., Oshima A.,
RA Sugiyama A., Kawakami B., Suzuki Y., Sugano S., Nagahari K.,
RA Masuho Y., Nagai K., Isegai T.;
RT "NEO human cDNA sequencing project.";
RL Submitted (MAR-2004) to the EMBL/GenBank/DBSJ databases.
DR EMBL; AK131434; BADI8580.1;
SQ SEQUENCE 933 AA; 102512 MW; B883773C5BC21A23 CRC64;

Query Match 54.4%; Score 3204.7; DB 2; Length 933;
Best Local Similarity 55.0%; Pred. No. 7.6e-54;
Matches 627; Conservative 0; Mismatches 0; Indels 513; Gaps 3;

QY 1 NAWKTLPIYLLLLSVFVIQVSSQDLSSCAGRCGEGYSRDTATCNCYHCQHYMECCPDF 60
Db 1 NAWKTLPIYLLLLSVFVIQVSSQ----- 25
QY 61 KRVTCTAELSCRGCFESPERGECDCDAQCCKYKCCPDYBSCAEVHNTPSPSSKAP 120
Db 26 -----EUSCKRGCFESPERGECDCDAQCCKYKCCPDYBSCAEVHNTPSPSSKAP 79
QY 121 PPGASQTIKSTTKRSPPPNKKTKKVIIESEEITE----- 156
Db 80 PPGASQTIKSTTKRSPPPNKKTKKVIIESEEITEHSVSENCESSSSSSSSSSTIR 139
QY 157 -----KVKONKNTKKKPTKPPVDEAGSLDNGDFKVTTPDTST 198
Db 140 KIKSKNSAANRELOKLVKNDKNTKKKPTKPPVDEAGSLDNGDFKVTTPDTST 199
QY 199 TQHNKVTSPKITTAKPINRPSLPNPSDTSKETSLSLVNKEITVETKETTITNKQSTDG 258
Db 200 TQHNKVTSPKITTAKPINRPSLPNPSDTSKETSLSLVNKEITVETKETTITNKQSTDG 259
QY 259 KEKTTSAKETOSIKTSKOLAPTSKVLAKPTPKAETTTKGPALTTPKPTTPPKEPAS 318
Db 260 KEKTTSAKETOSIKTSKOLAPTSKVLAKPTPKAETTTKGPALTTPKPTTPPKEPAS 319

```

RESULT 5

```

Q9JNM99 ID Q9JNM99 PRELIMINARY; PRT; 1054 AA.
AC Q9JNM99;
DT 01-OCT-2000 (TREMELrel. 15, Created)
DT 01-OCT-2000 (TREMELrel. 15, Last sequence update)
DT 01-MAR-2004 (TREMELrel. 26, Last annotation update)
DE Mus musculus.
GN Name=Prg4;
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=20573856; PubMed=11124536;
RT Ikegawa S., Sano M., Koshizuka Y., Nakamura Y.;
RT "Isolation, characterization and mapping of the mouse and human PRG4
RT (proteoglycan 4) genes.";

```

```

RL Cytogenet. Cell Genet. 90:291-297(2000).
DR EMBL; AB034730; BAA92310.1; -.
DR HSSP; PQ4004; I0C0.
DR MGD; MGI:1891344; Prg4.
DR InterPro; IPR000585; Hemopexin.
DR InterPro; IPR001212; Somatomedin_B.
DR Pfam; PF00045; Hemopexin; 2.
DR Pfam; PF01033; Somatomedin_B; 2.
DR PRINTS; PR00022; SOMATOMEDIN.
DR SMART; SM00120; HX; 2.
DR SMART; SM00201; SO; 2.
DR PROSITE; PS00024; HEMOPEXIN; UNKNOWN_1.
DR PROSITE; PS00524; SOMATOMEDIN_B; 2.
SQ SEQUENCE 1054 AA; 115991 MW; 4FC64BFA42283235 CRC64;

Query Match          46.4%; Score 2731.5; DB 2; Length 1054;
Best Local Similarity 47.1%; Pred. No. 9.3e-45;
Matches 545; Conservative 43; Mismatches 143; Indels 425; Gaps 21;

QY 1 MAWKTLPIYILLLSVFIQVSSQDLSSCAGRCGEGYSDATCNCYNCQHYMCCPDF 60
Db 1 MGWKILPVCLSLLPVLIQVSSQDLSSCAGRCGEGYSDATCNCYNCQHYMCCPDF 60

QY 61 KRVCTAELCKGRCFSPERGECDDAOKKYDKCCPDYSEFCABVHNPTSPSKKAP 120
Db 61 KRVCSPELCKGRCFSPERGECDDSQCKYKCCADYDFCEVHNSTSPSKTAP 119

QY 121 PPSGASQTIKSTTKRKPKNKKTKKV7ESEITEK----- 157
Db 120 TPAGASDTIKSTTKRSPKFT-TRTIKVSESELTEHSDSENQSSSSSSSSSTIRKI 178

QY 158 -----VKONKNRTKKKTPPPVVDVDEAGSLDNGDFKT--TDTSTT 199
Db 179 KSSNSANRELQXNVNKKNTKPKENPEPPAVDEAGSLDNGDFKLTTPPDPTT 238

QY 200 QENKSTSKITTAKEINPRSLPPNSDTSKSTSLVNKETVENKETTNNKQSTGOK 259
Db 239 PSKVAISPTTAKEVTPSPSLAPNSETSKASLASNKETIVETKETTATNKQSA-SK 297

QY 260 EKTTSAKETQSIEKTSKVLAKPTPKAETTTKGPALTTPKEPTTPKEPAT 319
Db 298 KKTTSVKETSAEKTSDKDVETS----- 321

QY 320 TPKEPTTIKSAPTTPKPAETTTKSAPTTKERAPTTKBPATTT-----PKBPA 371
Db 322 -----TTPKNSAPTTTKKPVTTTKESKFLPQBPPE 352

QY 372 PTTKEPATTTKSAPTTPKPAETTPKPAETTPKPAETTPKPAETTPKPAETTK 431
Db 353 PTAKEPPTTKKPEPTTKPEPTTKPEPTTKPEPTTKPEPTTKPEPTTKK 412

QY 432 PAPTTPKEPATPKAPKAPKAPKAPKAPKAPKAPKAPKAPKAPKAPKAPKAP 491
Db 413 PEPTTPKEPGPTTPKEPEPTTKKEPEPTTKKEPEPTTKKEPEPTTKKEPE 461

QY 492 TKEPATTTKSAPTTKKESPTTKKPAETTPKPAETTPKPAETTPKPAETTK 551
Db 462 -----EPPTTKKEPEPTTKKPEPTTKKPEPTTKKPEPTTKKPEPTTK 510

QY 552 PTTTKKPAETPKAPETPKETAPTTTPKXLPFTTPEKLPATTPKPAETTP 611
Db 511 PT----- 512

QY 612 EEPPTTPKEPATTPKAAAPTTPKPAETTPKPAETTPKPAETTPKPAETTP 671
Db 513 -----TPKEPEPTTKKPEPTTKKPEPTTKKPEPTTKKPEPTTK 538

QY 672 PTTLKEPATTPKPKAPKAPKAPKAPKAPKAPKAPKAPKAPKAPKAPKAP 731
Db 539 ----KEPEPTTKKP-----EPPTTKKEPVPTTKPEP 565

QY 732 APPTPKGAPTTLKEPATTPKPKAPKAPKAPKAPKAPKAPKAPKAPKAPKAP 791

```

```

Db 566 EPPTP-----KEPEPTTKPEPE-----TTRKEPEPTTKPEPEPTTKPE 604
QY 792 PAPTTPKAPAPTTPETPTTSEVSTTTTKEPTTIHKSPDESTPELSAETPKALENSP 851
Db 605 PEPTTPKKEPEPT----- 617
QY 852 KPEGVPTTKTPAATKPEMTTAKDKTTERDLRTTPTTAAAPKMTKETATTTKTESKI 911
Db 618 -----TSFKT----- 622
QY 912 TATTQVSTTTQDTTPFKITLTTTLAPKVTTKTKTTTTEIMNKPEETAKPKDRATN 971
Db 623 -----TTLKATTLAPKVTAPAE-----ETQNKPEETTPASESDSD 657
QY 972 SKAT-----TPKP-QKPTKAPKPTSTKPKTPRVRKPKTTPPRKMTSTMPSLNP 1022
Db 658 SKTLKPKQPKAPKPKPKPKAPKPKPKPKPKPKPKPKPKPKPKPKPKPKPKPK 716
QY 1023 TSRIAEAMLQTTTRPNQTPNSKLVNPKSEDAGAGETPHMLLRPHVFMPEYTPDMY 1082
Db 717 TP--LEVMLPTTTPKQTPNPETAENVNPDHEDADGEGEKP-LIPGPPVLPFTAIPGTDL 773
QY 1083 LPRVFNQGIINPMLS 1098
Db 774 LAGRLNRGININPMPS 789

RESULT 6
MUC2_HUMAN
ID MUC2_HUMAN STANDARD; PRT; 5179 AA.
AC Q02817; Q14878;
DT 01-JUN-1994 (Rel. 29, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 05-JUL-2004 (Rel. 44, Last annotation update)
DE Mucin 2 precursor (Intestinal mucin 2).
GN Name=MUC2; Synonyms=SMUC;
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Intestine;
RX MEDLINE=94132002; PubMed=8300571;
RA Gum J.R. Jr., Hicks J.W., Toribara N.W., Siddiki B., Kim Y.S.;
RT "Molecular cloning of human intestinal mucin (MUC2) cDNA.
RT Identification of the amino terminus and overall sequence similarity
RT to prepro-von Willebrand factor.";
RL J. Biol. Chem. 269:2440-2446(1994).
RN [2]
RP SEQUENCE OF 626-1895 AND 4196-5179 FROM N.A.
RC TISSUE=Colon;
RX MEDLINE=93016075; PubMed=1400449;
RA Gum J.R. Jr., Hicks J.W., Toribara N.W., Rothe E.-M., Lagace R.E.,
RA Kim Y.S.;
RT "The human MUC2 intestinal mucin has cysteine-rich subdomains located
RT both upstream and downstream of its central repetitive region.";
RL J. Biol. Chem. 267:21375-21383(1992).
RN [3]
RP SEQUENCE OF 1343-1895 AND 4176-4195 FROM N.A.
RX MEDLINE=91358717; PubMed=1885763;
RA Toribara N.W., Gum J.R. Jr., Culhane P.J., Lagace R.E., Hicks J.W.,
RA Petersen G.M., Kim Y.S.;
RT "MUC-2 human small intestinal mucin gene structure. Repeated arrays
RT and polymorphism.";
RL J. Clin. Invest. 88:1005-1013(1991).
CC -!- FUNCTION: Coats the epithelia of the intestines, airways, and
CC other mucous membrane-containing organs. Thought to provide a
CC protective, lubricating barrier against particles and infectious
CC agents at mucosal surfaces.
CC -!- SUBUNIT: Multimeric.
CC -!- SUBCELLULAR LOCATION: Secreted.
CC -!- TISSUE SPECIFICITY: Colon, small intestine, colonic tumors,

```

CC bronchus, cervix and gall bladder.
 CC -!- PTM: All cysteine residues are involved in intrachain or
 CC interchain disulfide bonds (By similarity).
 CC -!- POLYMORPHISM: The number of repeats is highly polymorphic and
 CC varies among different alleles.
 CC -!- SIMILARITY: Contains 1 C-terminal cysteine knot-like (CTCK) domain.
 CC -!- SIMILARITY: Contains 1 TIL (Typein inhibitory-like) domain.
 CC -!- SIMILARITY: Contains 2 WFEC domains.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; L21998; AAB95295.1; -
 CC EMBL; M74027; AAB59875.1; -
 CC EMBL; M94131; AAB59163.1; -
 CC EMBL; M94132; AAB59164.1; -
 CC PIR; A49963; A43932.
 CC Genew; HGNC:7512; MUC2.
 CC MIM; 158370; -
 CC InterPro: IPR002919; Cysrich TIL.
 CC InterPro: IPR006208; Cys knot.
 CC InterPro: IPR006207; Cys_knot_C.
 CC InterPro: IPR006209; EGF_like.
 CC InterPro: IPR001007; VWF_C.
 CC InterPro: IPR001846; VWF_D.
 CC Pfam; PF00007; Cys_knot; 1.
 CC Pfam; PF01826; TIL; 1.
 CC Pfam; PF00093; VMC; 1.
 CC Pfam; PF00094; VMD; 4.
 CC SMART; SM00041; CT; 1.
 CC SMART; SM00214; VMC; 2.
 CC SMART; SM00216; VMD; 4.
 CC PROSITE; PS01185; CTCK_1; 1.
 CC PROSITE; PS01225; CTCK_2; 1.
 CC PROSITE; PS00022; EGF_1; UNKNOWN_1.
 CC PROSITE; PS01208; VMC_1; 2.
 CC PROSITE; PS0184; VMC_2; 2.
 KW Glycoprotein; Repeat; Signal.
 FT SIGNAL 1 20 Potential.
 FT CHAIN 21 5179 Mucin 2.
 FT DOMAIN 1401 1747 APPROXIMATE REPEATS.
 FT REPEAT 1401 1416 1.
 FT REPEAT 1417 1432 2.
 FT REPEAT 1433 1448 3.
 FT REPEAT 1449 1464 4.
 FT REPEAT 1465 1471 5.
 FT REPEAT 1472 1478 6.
 FT REPEAT 1479 1494 7A.
 FT REPEAT 1495 1517 7B.
 FT REPEAT 1518 1533 8A.
 FT REPEAT 1534 1556 8B.
 FT REPEAT 1557 1572 9A.
 FT REPEAT 1573 1596 9B.
 FT REPEAT 1597 1612 10A.
 FT REPEAT 1613 1635 10B.
 FT REPEAT 1636 1651 11A.
 FT REPEAT 1652 1675 11B.
 FT REPEAT 1676 1693 12.
 FT REPEAT 1684 1699 13.
 FT REPEAT 1700 1715 14.
 FT REPEAT 1716 1731 15.
 FT REPEAT 1732 1747 16.
 FT DOMAIN 4815 4886 VWF_C_1.
 FT DOMAIN 4924 4991 VWF_C_2.
 FT DOMAIN 5075 5160 CTCK.
 FT DISULFID 5075 5122 By similarity.
 FT DISULFID 5089 5136 By similarity.
 FT DISULFID 5089 5152 By similarity.

FT DISULFID 5102 5154 By similarity.
 FT DISULFID 5159 5159 By similarity.
 FT CARBOHYD 163 163 N-linked (GlcNAc...) (Potential).
 FT CARBOHYD 423 423 N-linked (GlcNAc...) (Potential).
 FT CARBOHYD 670 670 N-linked (GlcNAc...) (Potential).
 FT CARBOHYD 770 770 N-linked (GlcNAc...) (Potential).
 FT CARBOHYD 894 894 N-linked (GlcNAc...) (Potential).
 FT CARBOHYD 1139 1139 N-linked (GlcNAc...) (Potential).
 FT CARBOHYD 1154 1154 N-linked (GlcNAc...) (Potential).
 FT CARBOHYD 1215 1215 N-linked (GlcNAc...) (Potential).
 FT CARBOHYD 1230 1230 N-linked (GlcNAc...) (Potential).
 FT CARBOHYD 1246 1246 N-linked (GlcNAc...) (Potential).
 FT CARBOHYD 1787 1787 N-linked (GlcNAc...) (Potential).
 FT CARBOHYD 1820 1820 N-linked (GlcNAc...) (Potential).
 FT CARBOHYD 4339 4339 N-linked (GlcNAc...) (Potential).
 FT CARBOHYD 4351 4351 N-linked (GlcNAc...) (Potential).
 FT CARBOHYD 4362 4362 N-linked (GlcNAc...) (Potential).
 FT CARBOHYD 4373 4373 N-linked (GlcNAc...) (Potential).
 FT CARBOHYD 4422 4422 N-linked (GlcNAc...) (Potential).
 FT CARBOHYD 4438 4438 N-linked (GlcNAc...) (Potential).
 FT CARBOHYD 4502 4502 N-linked (GlcNAc...) (Potential).
 FT CARBOHYD 4615 4615 N-linked (GlcNAc...) (Potential).
 FT CARBOHYD 4627 4627 N-linked (GlcNAc...) (Potential).
 FT CARBOHYD 4752 4752 N-linked (GlcNAc...) (Potential).
 FT CARBOHYD 4787 4787 N-linked (GlcNAc...) (Potential).
 FT CARBOHYD 4881 4881 N-linked (GlcNAc...) (Potential).
 FT CARBOHYD 4888 4888 N-linked (GlcNAc...) (Potential).
 FT CARBOHYD 4955 4955 N-linked (GlcNAc...) (Potential).
 FT CARBOHYD 4970 4970 N-linked (GlcNAc...) (Potential).
 FT CARBOHYD 5019 5019 N-linked (GlcNAc...) (Potential).
 FT CARBOHYD 5038 5038 N-linked (GlcNAc...) (Potential).
 FT CARBOHYD 5069 5069 N-linked (GlcNAc...) (Potential).
 FT CONFLICT 1351 1351 H -> L (in Ref. 3).
 FT CONFLICT 1412 1412 T -> S (in Ref. 3).
 FT CONFLICT 1449 1449 L -> P (in Ref. 3).
 FT CONFLICT 1504 1504 M -> T (in Ref. 3).
 FT CONFLICT 4192 4192 G -> S (in Ref. 2).
 SQ SEQUENCE 5179 AA; 540295 MW; 85CD7571FB9A5663 CRC64;
 Query Match 23.7%; Score 1396.9; DB 1; Length 5179;
 Best Local Similarity 20.1%; Pred. No. 1.9e-18;
 Matches 449; Conservative 85; Mismatches 494; Indels 1211; Gaps 71;
 QY 6 LPVILLLLSVFIQVSSQDLSSCARGC-----EG-----YSR 40
 DB 502 LQVLAPVWQLFVTLDOAQ--GQVGLCGNFNGLEGDFKTASGLVEATGAGFANTWKA 559
 QY 41 DATCN-----CDY---- 48
 DB 560 QSTCHDKLDLDDPCSLINIESANYAEHWCSSLKKTETPGRCHSAVDPAEYVKRKYDTG 619
 QY 49 NCOHYMEC-----CPDFKRVCTA-----ELSK 71
 DB 620 NQNNEDCLCAALSVARACTAKGVMLGWREHVCNKGVCSPNSQVFLNYLTTCQOTCR 679
 QY 72 -----GRCPESP-----ERGR-----ECDC-----DAQCKYDK 95
 DB 680 SLSEADSHCLEGAPVDPVGGCCPDHTFLDEKRCVFLAKSCYHRGLYLAGDVVVEER 739
 QY 96 C-----YESFCAEVHN----- 109
 DB 740 CVCRDRLHCRQIRLIGQSCCTAPKIHMDCSNLTALATSKPRALSCOTLAAGYHTECVSG 799
 QY 97 --CPD-----YESFCAEVHN----- 109
 DB 800 CVPDGLMDDRGCGGVVEKECPCHVHNDLYSSGAKIKVDNCTCTCKRGWVCTQAVCHGT 859
 QY 110 ----- 109
 DB 860 CSYVSGSHYITFDGKYVDFDGHGSYVAVQDYCQNSLGLSFSIITENVPGTGTGVTCSKA 919
 QY 110 ----- 109

Db 920 IKIFMGSTELKLEDRVVIQRDEGHVAYTTREVCGQYLVBESSGCIIVINDKRTTVFIK 979
QY 110 -----PTSPSSKKAPPS- 123
Db 980 LAPSXYGTVCGLGCFDHRNNDFTTRDHVWVSSELDGNSWKEAPCTPDVSTNPECSL 1039
QY 124 -----GASQTIKST-----TKESKP----- 139
Db 1040 NPHRRSWAEKQCSILKSVFSICHKVDKPFYACVHSCDSCDGGDCBCEFCSCASAVA 1099
QY 140 ----- 139
Db 1100 QECTKEGACVFWRTPLDCPIFCDYNNPPECEWHYBPCGNRSFETCORTINGIHSNLSVSY 1159
QY 140 -----PNKKKTKVIESBEITEKVDNKKNRKKK-----TPPKPP----- 175
Db 1160 LEGYPCRPADRIYE-----EDLKKCVTADKCGCVEDTHYPPGASVTEBCKSC 1211
QY 176 -----VDEAGSGL-----DNG----- 187
Db 1212 VCTNSSQVCRPEEGKILNQTDGAFYWEICGPNGTVEKHFNCISITRPSLTITFTI 1271
QY 188 -----DFKVTPTDSTQHKNKVTSPK----- 209
Db 1272 TUPPTPTSTTTTTTPTSTSTVLSTPKLCLWSDWINEHPSGSDGDRPFDGVC 1331
QY 210 ----- 209
Db 1332 APEDIECRSVKPHLSLEOHGQKQCDVSUGVICKNEDQFNGPGLCYDYKIRVNCWP 1391
QY 210 -----INTAKPINRPSLPNSDTSKETSITVKNKETTVEKETTITNNKQSTDGKSKTISA 265
Db 1392 MDKCIITPSPPPTTTPSPPTTTTTPPTTTPSPPT-----TITTPPTTTPSPPTT 1446
QY 266 KETQSTKTSKADLAPTSKVLAKPTKAEITTKGPAITTKPEPT-----PTPKPASTT 320
Db 1447 TP-----TPSPISITTTTPPTTTPSPPTTTPSPPTTTPSPPTTTPSPPT 1490
QY 321 PREFTPTTKSAP-TPPKP-----APTTKSAP-TPPKPAPTTPKAPAPTTPKBP- 370
Db 1491 TTTTPPTTTPSPPTTTPASTTTTTPPTTTPSPPTTTPPTTTPPTTTPSPPTTTP 1550
QY 371 -----APTTPKAPAPTTPKAP-TPPKPAPTTPKAPAPTTPKAPAPTTPKAPTTP 422
Db 1551 TSTTLTPPTTTPSPPTTTPPTTTPSPPTTTPPTTTPPTTTPPTTTPPTTTPPT 1610
QY 423 KEPAPTTPKAPAPTTPKAP-----APTAPKAPAPTTPKAPAPTTPKAPAPTTPKBP 475
Db 1611 TPTPTTTPSPPTTTPPTTTPPTTTPPTTTPPTTTPPTTTPPTTTPPTTTPPT 1670
QY 476 TPKEPAPTTPKAPAPTTPKAPAPTTPKAP-----TPPKPAPTTPKAPAPTTPKAPAPTTPK 532
Db 1671 TTTTPPTTTPSPPTTTPSPPTTTPPTTTPPTTTPPTTTPPTTTPPTTTPPTT 1729
QY 533 KPAPTTPKAPAPTTPKAPAPTTPKAPAPTTPKAPAPTTP-----KETAPTTPKLTTP--- 584
Db 1730 TSPPTTTPSPPTTTPPTTTPPTTTPPTTTPPTTTPPTTTPPTTTPPTTTPPT 1787
QY 585 ----- 584
Db 1788 WTGWLDSGKPNFKPGDTELIGDVGCGWAANISCRATMYPDVPIGQLGQTVWCDVSG 1847
QY 585 -----TPPEKLAPTTPKAPAPTTPPELAPTTP 612
Db 1848 LICKNEDQKPGVIMAFCLNVEINVQCECVTQPTTMTTNTTTPPTTTPPTTTP 1907
QY 613 EPTP-----TPPEAPT-----TPKAAAPTTP-- 635
Db 1908 TTPPTPTGQTPTTPTTPTTPTTPTGQTPTTPTTPTTPTTPTGQTPTTPTGQTPTT 1967
QY 636 -----KEPAPTTPKAP-----APTTPKAPAPTTPKAPAPTTPKAPAPTTPKAPA 679
Db 1968 TPTTPTTPTPTPTPTPTPTPTTPTTPTTPTTPTTPTPTPTPTPTPTPTPTPTPT 2027

QY 680 PTTPKKAPKELAPTTPKAP-----TSTTSKAPPT----- 710
Db 2028 PTGQTPTTPTTPTTPTTPTTPTGQTPTTPTTPTTPTTPTTPTTPTTPTTPTT 2087
QY 711 -----TPKGT-----APTTPKAP-----TPPKAPPT----- 734
Db 2088 TTTVTPPTPTGQTPTTPTTPTTPTTPTTPTGQTPTTPTTPTTPTTPTTPTTPT 2147
QY 735 -----TPKGT-APT-----LKEPAPTTPKAPKAPKELAPTTPK 767
Db 2148 TPTTPTTPTTPTTPTTPTGQTPTTPTTPTTPTTPTTPTTPTTPTTPTTPTTPT 2207
QY 768 P-----TSTTSKAP-----TPPKAPTTPKAP----- 794
Db 2208 PTPPTGQTPTTPTTPTTPTTPTTPTGQTPTTPTTPTTPTTPTTPTTPTTPTT 2267
QY 795 -----TPPKKAPPT-----TPETP-----PTTSVPS-----TP 818
Db 2268 PTTTPTTPTTPTTPTGQTPTTPTTPTTPTTPTTPTTPTTPTTPTTPTTPTTPT 2327
QY 819 TTTKEPTTIHKSPESTPELSAETPKALENSPKP-----GVPT-TKTPAATKP 867
Db 2328 TGTQPTT-----TPITTTTPTTPTTPTTPTTPTTPTTPTTPTTPTTPTTPTT 2384
QY 868 EMITTAKDKITERDLR---TTP---ETTTAAPKMT---KETATTTKTESKI----- 911
Db 2385 TTTTPTTPTTPTGQTPTTPTTPTTPTTPTTPTTPTTPTTPTTPTTPTTPTTPT 2444
QY 912 -TATTQVSTT-----TQDTPPKITT-----LKTT 938
Db 2445 TQTPPTTPTTPTTPTTPTTPTTPTTPTTPTTPTTPTTPTTPTTPTTPTTPTT 2504
QY 939 LAPVTTT-KKTIITTIMKPEETAKPKORATNSKATP-----KPKPT 983
Db 2505 VTPPTPTGQTPTTPTTPTTPTTPTTPTTPTTPTTPTTPTTPTTPTTPTTPTTPT 2564
QY 984 KAP-----KKPTSTKKKIMPRVKPKITTPPKMTSTMPELNPTSRIAEAMLOT 1033
Db 2565 TPTTPTTPTTPTTPTGQTPTTPTTPTTPTTPTTPTTPTTPTTPTTPTTPTTPT 2616
QY 1034 TT-----RPNQTPNSKLVEVNPKSEDAG-----GAE 1059
Db 2617 TTTVTPPTGQTPTTPTTPTTPTTPTTPTTPTTPTTPTTPTTPTTPTTPTTPT 2676
QY 1060 GETPHMLLRHVFMEVPTP 1078
Db 2677 TPTTPTTPTTPTTPTTPT 2695

RESULT 7

Q9VR49
ID Q9VR49 PRELIMINARY; PRT; 1225 AA.
AC Q9VR49;
DC 01-MAY-2000 (TREMBlrel. 13, Created)
DT 01-JUN-2003 (TREMBlrel. 24, Last sequence update)
DT 01-MAR-2004 (TREMBlrel. 26, Last annotation update)
DE CG3047-PA.
GN Name=Sgs1; ORFNames=CG3047;
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
RX NCBI_taxid=7227;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA George R.A., Lewis S.B., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.H., Blaise R.G., Champe M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Gabor G.L.,

RA Abril J.F., Agbavani A., An H.J., Andrews-Pfannkuch C., Baldwin D.,
 RA Ballaw R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
 RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
 RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
 RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
 RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
 RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
 RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
 RA Durbin K.J., Evangelista C.C., Ferraz C., Gelbart W.M., Gleason K.,
 RA Foster C., Gabrielson A.E., Garg N.S., Gelbart W.M., Gleason K.,
 RA Glodok A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
 RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
 RA Hostin D., Houston K.A., Howland T.J., Wei M.H., Ibegwam C.,
 RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.W., Ketchum K.A.,
 RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
 RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
 RA Liu X., Mattle B., McIntosh T.C., McLeod M.P., McPherson D.,
 RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
 RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
 RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
 RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
 RA Reinert K., Remington K., Saunders R.D., Scheeler F., Shen H.,
 RA Shue B.C., Siden-Kianos I., Simpson M., Skupski M.P., Smith T.,
 RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
 RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
 RA Wang Z.Y., Wasserman D.A., Weinstein G.M., Weissenbach J.,
 RA Williams S.M., Woodgett, Worley K.C., Wu D., Yang S., Yao Q.A., Ye J.,
 RA Yeh R.F., Zaveri J.S., Zhan M., Zhang K., Zhao Q., Zheng L.,
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.,
 RT "The genome sequence of *Drosophila melanogaster*."
 RL Science 287:2185-2195(2000).
 RN [2]
 RN SEQUENCE FROM N.A.
 RX MEDLINE=22426065; PubMed=12537568;
 RA Celisner S.E., Wheeler D.A., Krommiller B., Carlson J.W., Halpern A.,
 RA Patel S., Adams M., Champe M., Dugan S.P., Frise E., Hodgson A.,
 RA George R.A., Hoskins K.A., Lavery T., Muzny D.M., Nelson C.R.,
 RA Pacleb J.M., Park S., Pfeiffer B.D., Richardson M., Sodergren E.J.,
 RA Svirskas R., Taber P.E., Wan K., Stapleton M., Sutton G.G., Venter C.,
 RA Weinstein G., Scherer S.E., Myers E.W., Gibbs R.A., Rubin G.M.;
 RT "Finishing a whole-genome shotgun: release 3 of the *Drosophila*
 RT melanogaster euchromatic genome sequence."
 RL Genome Biol. 3:RSEARCH0079-RESEARCH0079(2002).
 RN [3]
 RN SEQUENCE FROM N.A.
 RX MEDLINE=22426070; PubMed=12537573;
 RA Kaminker J.S., Bergman C.M., Krommiller B., Carlson J., Svirskas R.,
 RA Patel S., Frise E., Wheeler D.A., Lewis S.E., Rubin G.M.,
 RA Ashburner M., Celisner S.E.;
 RT "The transposable elements of the *Drosophila melanogaster* eu-
 RT chromatin: a genomic perspective."
 RL Genome Biol. 3:RSEARCH0084-RESEARCH0084(2002).
 RN [4]
 RN SEQUENCE FROM N.A.
 RX MEDLINE=22426069; PubMed=12537572;
 RA Mitra S., Crosby M.A., Mungall C.J., Matthews B.B., Campbell K.S.,
 RA Hradecky P., Huang Y., Kaminker J.S., Millburn G.H., Prochnik S.E.,
 RA Smith C.D., Tupy J.J., Whitfield E.J., Bayraktaroglu L., Berman B.P.,
 RA Bettencourt B.R., Celisner S.E., de Grey A.D., Drysdale R.A.,
 RA Harris N.L., Richter J., Russo S., Schroeder A.J., Shu S.Q.,
 RA Stapleton M., Yamada C., Ashburner M., Gelbart W.M., Rubin G.M.,
 RA Lewis S.E.;
 RT "Annotation of the *Drosophila melanogaster* euchromatic genome: a
 RT systematic review."
 RL Genome Biol. 3:RSEARCH0083-RESEARCH0083(2002).
 RN [5]
 RN SEQUENCE FROM N.A.
 RP FLYBASE;
 RG Submitted (SEP-2002) to the EMBL/GenBank/DBJ databases.
 RN [6]
 RN SEQUENCE FROM N.A.
 RP FLYBASE;

RL Submitted (MAR-2004) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AB003575; AAF50957.3; --
 DR FLYBASE: FBgn0003372; Sgsl;
 SQ SEQUENCE 1225 AA; 127735 MW; A99AF9D4404C79F4 CRC64;
 Query Match 23.0%; Score 1352.9; DB 2; Length 1225;
 Best Local Similarity 27.5%; Pred. No. 2.2e-18;
 Matches 366; Conservative 121; Mismatches 422; Indels 421; Gaps 53;
 QY 10 LLLLSVFIQV-----SSQDLSSCAGCGEGSRDATCNCYNCQHYMECCDDF 60
 DB 6 IFUTVSILLIQVKNVANDWSMD-----GFSEIIPGCGGDIY-----PD- 50
 QY 61 KRYCTAELSKGRCFESFERGECDCDAOCKYKCCPDYESFCBAEVHNPTSPSSKAP 120
 DB 51 -----PVQPCDTDS-----NPTTKRQKTK 71
 QY 121 PRGASQTIKSTKSPKPPNKKYKVEISEITEKYKDKNKKKKP-TPKEPVVDE 179
 DB 72 RP-----KSTRRTKRT-KRPRKTKWTTK-----RATKRTTKRTRRRPTTKTP- 117
 QY 180 AGSGLDNGDFKVTTPDT-----STQHNKVS-----TSPKITTAKPINRPS 221
 DB 118 -----DTTDSPIITGAECTCSDRITASDSDSTDRITVNTDWTPLCTDPEPCT- 168
 QY 222 LPNSDTSKETSITVNNKTTVT-----KEITTNKQTSDDGKEK--TSAKTQ 269
 DB 169 -----CSSESSTAIPISSPCIDTSTVIPTSCQBTETPTPTCSQGTQTTCTCAQTT- 221
 QY 270 SIKTSAKDLAPTSKVLAKPTPKAETT--TKGPALTTPKEPTTPKGPASTTPKEPTPT 327
 DB 222 -----TTPRSITTTSTSPRTTTTPRSITTTTTSRPTTTTTPRSTTTT 262
 QY 328 TIKSAPTTKPEAPTTTKSAP--TTPKEAPTTTKPEAPTTTKPEAPTTTKPEAPTTTKS 385
 DB 263 TTRRPTTTTTPRCTTTTSCAPTTPRSTTTTTSRPTTTTTPRCTTTTSCPTTTPRS 322
 QY 386 APPT-PKEAPPTP-----KKPAPTTPK-----PAPTTPK-----P 417
 DB 323 TTTTTSRPTTTTPRCTTTTPSTTCTTRPTTTTPRSTTTTCTSGPTTTTPRSTTTTCTSGP 382
 QY 418 TPTTPKEAPT-TKEPAPTTKPEAPTPAPKAPTTPK-----PAPTTKPEAPT 468
 DB 383 TTTTTPRSTTTTSGPTTTTPRSTTTTSGPTTTTPRSTTTTSGPTTTTTPRSTTTT 442
 QY 469 TKEPSTTPKEAPTTTKSAPTTKPEAPTTTKSAP--TTPKEPSTTTKPEAPTTKPE- 525
 DB 443 TSGPTTTTTPRSTTTTSGPTTTTPRSTTTTSGPTTTTTPRSTTTTSGPTTTTTPRST 502
 QY 526 -----PAPTTPK-----PAPTTPK-----PAPTTKPEAPTTTKPAPTA 562
 DB 503 TTTTSGPTTTTPRSTTTTSGPTTTTPRSTTTTSGPTTTTTPRSTTTTSGPTTTT 562
 QY 563 PKE-----PAPTTKETAPTTPKLTPTTPEKLAPTTPKAPTTPEELAPTTPEP 614
 DB 563 PRSTTTTSGPTTTTTPRSTTTTSGPTTTTTPRSTTTTSGPTTTTTPRSTTTTSGP 622
 QY 615 TPTTPEE-----PAPTTPKAAAPNTKPEAPTTPK-----PAPTTPK----- 653
 DB 623 TTTTTPRSTTTTSGPTTTTTPRSTTTTSGPTTTTTPRSTTTTSGPTTTTTPRSTTTT 682
 QY 654 ---PAPTTPKETAP-----TTPKGATPTLKGPAPTTPK-----PAPKEL 691
 DB 683 TSGPTTTTTPRSTTTTSGPTTTTTPRSTTTTSGPTTTTTPRSTTTTSGPTTTTTPRST 742
 QY 692 APPTTKEPTSTT-----SDKAPPTTPKGAPTTKPEAPTTPK-----PAPT 735
 DB 743 TTTTSGPTTTTPRSTTTTSGPTTTTTPRSTTTTSGPTTTTTPRSTTTTSGPTTTT 802
 QY 736 PKGTAPTTLKPEAPTTPK-----PAPKELAPTTTKGPTSTT-----SDKP 776
 DB 803 PRSTTTTSGPTTTTTPRSTTTTSGPTTTTTPRSTTTTSGPTTTTTPRSTTTTSGP 862

```

QY 777 APTPKETAPTKPAPTKPKK-----PAPTKPET-----PPPTTSEVSTPTT 820
DB 863 TTTTPRSTTTCTSGPTTTTPRSTTTCTSCPTTTTPRSTTTCTSCPTTTTPRSTTTTC 922
QY 821 TKEPTTIHKSPDESPELSAETPRALENSPEKPGVPTKPAAPKPEMTTAKDKITR 880
DB 923 TSGPTT--TTPRSTTKTSCAPT-----TTPRSTTTTTSRPTTTTPRSTTTT--TTSR 973
QY 881 DURTPTETTAAPKMKETATTEKTBESKIPATTOVTSTTQDTTTPFKITLTKTTLA 940
DB 974 PTTTTPRSTT--TPSTRPTTTTPRSTT-----TTSRPTTTTPRSTT-----KTSICA 1021
QY 941 PKVTITTKTITTEIMNRPEETAKPKDRATNSKATPKPKETKAPK-----KPTST 992
DB 1022 PTTTTPRSTTTT-----TSRPTTTTPRSTTTTTSRPTTTTPRSTTTTPCTSRPTTT 1073
QY 993 KKPWKMPRVKPKTTPPKMT-----STWPELNPSTRIAEAMLOTTTR-----1036
DB 1074 TPRSITTTTTSRPTTTTPRSTTTPCPTTPSPR-----TTPTRPCPCHPQPYQ 1126
QY 1037 -----PNQT-----PNSKLVEVNPKSEDAGGAGETPHMLRPHVFMPEVTPD 1079
DB 1127 IPPWSWVNPNTYPNPFVWFQPN-----PVPFQ 1154
QY 1080 MDYLPVRVNO 1089
DB 1155 WPMQGYNO 1164

RESULT 8
ID Q7KTF6 PRELIMINARY; PRT; 1761 AA.
AC Q7KTF6;
CT 05-JUL-2004 (TReMBLrel. 27, Created)
DT 05-JUL-2004 (TReMBLrel. 27, Last sequence update)
DE CG33300-PA (Fragment).
GN ORFNames=CG33300;
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.H., Blazek R.G., Champe M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Gabor G.L.,
RA Abril J.F., Agbayani A., An H.J., Andrews-Pfannkoch C., Baldwin D.,
RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brotter P.,
RA Burtis K.C., Busan D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
RA Foster C., Garg N.S., Guan P., Guan P., Harris M.,
RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harsis M.,
RA Harris N.L., Harvey D., Helman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.H., Ibegwam C.,
RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Maitel B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacle B.J.M.,
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,

```

```

RA Reinert K., Remington K., Saunders R.D., Scheeler F., Shen H.,
RA Shue B.C., Siden-Kamos I., Simpson M., Skupski M.P., Smith T.,
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.Y., Wassarman D.A., Weinstein G.M., Weissenbach J.,
RA Williams S.M., Woodage, Worley K.C., Wu D., Yang S., Yao Q.A., Ye J.,
RA Yeh R.F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RA "The genome sequence of Drosophila melanogaster.";
RL Science 287:2185-2195(2000).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=22426065; PubMed=12537568;
RA Celniker S.E., Wheeler D.A., Kronmiller B., Carlson J.W., Halpern A.,
RA Patel S., Adams M., Champe M., Dugan S.P., Frise E., Hodgson A.,
RA George R.A., Hoskins R.A., Javerty T., Muzny D.M., Nelson C.R.,
RA Pacle B.J.M., Park S., Pfeiffer B.D., Richards S., Sodergren E.J.,
RA Svirskas R., Tabor P.E., Wan K., Stapleton M., Sutton G.G., Venter C.,
RA Weinstein G., Scherer S.E., Myers E.W., Gibbs R.A., Rubin G.M.;
RA "Finishing a whole-genome shotgun: release 3 of the Drosophila
melanogaster euchromatic genome sequence.";
RL Genome Biol. 3:RESEARCH0079-RESEARCH0079(2002).
RN [3]
RP SEQUENCE FROM N.A.
RX MEDLINE=22426070; PubMed=12537573;
RA Kaminker J.S., Bergman C.M., Kronmiller B., Carlson J., Svirskas R.,
RA Patel S., Frise E., Wheeler D.A., Lewis S.E., Rubin G.M.,
RA Ashburner M., Celniker S.E.;
RA "The transposable elements of the Drosophila melanogaster euchromatin:
a genomics perspective.";
RL Genome Biol. 3:RESEARCH0084-RESEARCH0084(2002).
RN [4]
RP SEQUENCE FROM N.A.
RX MEDLINE=22426069; PubMed=12537572;
RA Misra S., Crosby M.A., Mungall C.J., Matthews B.B., Campbell K.S.,
RA Hradecky P., Huang Y., Kaminker J.S., Millburn G.H., Prochnik S.E.,
RA Smith C.D., Tupy J.L., Whitfield E.J., Bayraktaroglu L., Berman B.P.,
RA Bettencourt B.R., Celniker S.E., de Grey A.D., Drysdale R.A.,
RA Harris N.L., Richter J., Russo S., Schroeder A.J., Shu S.Q.,
RA Stapleton M., Yamada C., Ashburner M., Gelbart W.M., Rubin G.M.,
RA Lewis S.E.;
RA "Annotation of the Drosophila melanogaster euchromatic genome: a
systematic review";
RL Genome Biol. 3:RESEARCH0083-RESEARCH0083(2002).
RN [5]
RP SEQUENCE FROM N.A.
RX FLYBASE;
RL Submitted (SEP-2002) to the EMBL/GenBank/DBJ databases.
RN [6]
RP SEQUENCE FROM N.A.
RX FLYBASE;
RL Submitted (MAR-2004) to the EMBL/GenBank/DBJ databases.
DR EMBL; AB003626; AAS64673.1; -
DR InterPro; IPR011009; Kinase_like.
FT NON TER 1
SQ SEQUENCE 1761 AA; 197521 MW; 62A5E16B6241B3F4 CRC64;
Query Match 22.5%; Score 1327.5; DB 2; Length 1761;
Best Local Similarity 24.7%; Pred. No. 1.1e-17;
Matches 451; Conservative 114; Mismatches 369; Indels 895; Gaps 85;
QY 74 CF-----ESFGRGECDCDAQCKKYDKCPDYSEFSC---104
DB 5 CFWLALYHSANNALEGSIKFGDSESTTSAE-----XY-----PWY---CWIN 45
QY 105 -----AVHNPTSPS 115
DB 46 PELPGCPFGGDNNSSTTKISTLATTKSTVTTSEETTLKITIKSTAKPTTKTNEPT 105
QY 116 SKKAPPSGA-----SQTIKST--TKR 135
DB 106 TEKITTPKATTIKSTATTATTAATEAPKTEQTTLRTTIKSTSELITLTKTTIKSTAEKTK 165

```


136 SPKPPNKKTKVIESEBIEKVDKNKKNKTKKPTKPPVVDGAGSLDNGDFKVT-- 193
166 STNPTTKKSTLRTTEPTTKSSTAK--TTREPTTK--RETTTER 207
194 ---PDTS--TTCNKNVSTSPKITTAKINRPPSLPPNSDTSKTSLVNKETIVE-- 243
208 TQEPSTSKTTHE---TAEPAATKTTHEPT--TQKSTLRTTEPTTKSSTA 257
244 --TKEITTNKQISDQKEKTSIAKTSQISIKTSKADLAPTSKVLAKPTPKAETTKGPA 301
258 KTTREPTTKRTTERTTKEPST--KTTTBE--TAEPAATKTTHEPTTQKSTI 307
302 LTTTKEPT-----PTTKPEAPATTKPEAPATTKPEA-----PT 342
308 LRTEPTTKSSTARTTRETTRTERTTQEPSTKTHETT--ABPATKTTHEPT 366
343 TTKSA-----PTTKSA-----PTTKPEAPATTKPKA----- 379
367 TQKSTLRTTEPTTKSSTAKTTRE--PTTKRTTERTTKEPSTSKTTTHTAABPATK 424
380 -----PTTKSA-----PTTKPEAPATTKPKA----- 402
425 KTTHEPTTQKSTLRTTEPTTKSSTARTTRETTRTERTTQEPSTKTHETTA 484
403 -----PTTKPEA-----PTTKPEPTTKPEAP-----PTT 436
485 EPATKTTHEPTTQKSTLRTTEPTTKSSTAKTTRETTRTERTTQEPSTSKTTT 544
437 PKPEAPATPKA--PTTK--EPA-----PTTKPEAPATTKEPS-- 473
545 HETTAETATKTTHEPTTQKSTLRTTEPTTKSSTARTTRETTRTERTTQEPST 604
474 -----PTTKPEA-----PTTKSA-----PT-- 490
605 SKTTTHTAEPATKTTHEPTTQKSTLRTTEPTTKSSTARTTRETTRTERTT 664
491 -----TTKPEA-----PTTKSA-----PTTKPEA-----PTTKPEP 510
665 QEPSTSKTTTHTAEPATKTTHEPTTQKSTLRTTEPTTKSSTAKTTRETTRT 724
511 SPTTKPEA-----PTTK--EPA-----PT-- 529
725 TERTTKEPTTKRTTTHKTTTTEPTTKKTTTTEPTTKKSTLTKPEPTTKRTKTTTTEPT 784
530 TPKEAPATTKPEA-----PTTKPEAPATTKPEA--PTAPK-----EPA 567
785 TKRTTERTTKEPTTKRTTTHKTTTTEPTTK--TTTKTTHTEPTTKKSTLTKPEPT 840
568 -----PTTKETA-----PTTKKLT----- 583
841 TRKSTTKTTEPTTKRTTERTTQEPSTSKTTTHTAEPATKTTHEPTTQKSTLRTI 900
584 ---PTT--PEKLAPTPEKPAPTTPELAPTPEPT-----PTTPE--EPA-- 623
901 TEEPTTKSSTAKTTRE---PTTKRETTERTKPEPTTKRTTTHKTTTTEPTTKTTHTEPT 957
624 -----PTTKAAAPTKEPAPTTPKPEAPATTKPEA-----PTTK 660
958 KKSSTLTKPEEPTTKRTTSTKTRE--PTTKRTTERTTKEPTTKRTTTHKTTTEPTTK 1015
661 ETA-----PTTKETA-----PTTKPEAPATTKPKAPKELA-----PTTKETS 701
1016 TTTKTTHTEPTTKSTLTKPEEPTTKRTTSTKTREPTTKRTTERTTRETPTTKTTH 1075
703 TTSKPA-----PTTKGTA-----PTTKPEAPATTKPEA-----PTTP 736
1076 ETAEPAKTTHEPTTQKSTLRTTEPTTKSSTAKTTRETTRTERTTKEPTTR 1135
737 KGTAPATTKPEA-----PTTK--KPAKELAPATTKGPTSTSDKPAPTPKETA 785
1136 KTTTHKTTTEPTTKTTHTEPTTKKSTLTKPEE---PTTK--TSTTKTTEPTTKRT 1190

786 PTTKPEA-----PTTK--KP-----APTTPE 807
1191 ERTTKEPTTKRTTTHKTTTEPTTKTTHTEPTTKSTLTKPEEPTTKRTTSTKT 1250
808 PPTTSEVSTPTTKPE-----TTIKSPDE-----STPELSAEPPTPKALENSKEPVP 857
1251 REPTTKRTTERTTRETPTTKTTHKTTTEPTTKTTHKTTTTEPTTK--STTLKPTBE 1309
858 TTKTAAAKPEMTTAKOTTERDLR-----TTPPE 887
1310 TTRKSTTKTRETPTTKRTTERTTRETPTTKTTHKTTTEPTTKTTHKTTTTEPTTK 1369
888 TTTAAP-----KTKETAT-----TTEKTESKITATTTQVSTTTQDTTPEKIT 932
1370 STTLKPEPTTKRTTSTKTREPTTKRTTERTTRETPTTKTTHKTTTTEPTTKT 1427
933 TTKT-----TTLAP-----KVTTK-----KTTTTEIMNKP-----EE 961
1428 TTKTTHTEPTTKSTLTKPEEPTTKRTTSTKTREPTTKRTTERTTRETPTTKTTHK 1487
962 TAKPKDRATNSKATPKQ-----KP-----TKAPKKPT----- 990
1488 TEEPTTKTTHKTTTTEPTTKSTLTKPEEPTTKRTTSTKTREPTTKRTTERTTRET 1547
991 -----STKKPKTMRVRKPTTTPRKTSTTTPRKTSTTTPRKTSTTTPRKTSTTTPR 1022
1548 TTKTTHKTTTTEPTTKTTHKTTTTEPTTKSTLTKPEEPTTKRTTSTKTREPTTK 1603
1023 TSRIEAMLOTT-----TRPNOTNSKLVEVNPKSEDAGAGET----- 1062
1604 TTRSTSVKTTADOTTKRTTAEKSTTQEPSTVSTTNSQSNSTTESITTEBQVHH 1663
1063 -----PHMLLRPH----- 1070
1664 HHHHHYHKPADLGSILPLPLPLPLPLPLPLPLPLPLPLPLPLPLPLPLPLPL 1723
1071 -----VFMPEVT--PDMDYLPRVN 1088
1724 PPLPEVNTAISLPEISLPLPLPLPLPLPLPLPLPLPLPLPLPLPLPLPLPL 1752
RESULT 9
AAS64673 PRELIMINARY; PRT; 1761 AA.
ID AAS64673 PRELIMINARY; PRT; 1761 AA.
AC AAS64673; (TREMELrel. 27, Created)
DT 01-APR-2004 (TREMELrel. 27, Last sequence update)
DT 01-APR-2004 (TREMELrel. 27, Last sequence update)
DE CG33300-PA (Fragment).
DN CG33300.
GN Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
CX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scherer S.E., Li F.W., Hoskins R.A., Galle R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers V.H., Blazek R.G., Champe M., Pfeiffer B.D.,
RA An J.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Gabor G.L.,
RA Abril J.F., Agayani A., An H.J., Andrews-Pfannkoeh C., Baldwin D.,
RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Bereman B.P., Bhandari D., Bolshakov S.,
RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brattier P.,
RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Dou P.L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,

RA Foster C., Gabriellian A.E., Garg N.S., Gelbart W.M., Glasser K., Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M., Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J., Hosain D., Houston K.A., Howland T.J., Wei M.H., Ibegwam C., Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A., Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z., Lesko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X., Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D., Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A., Mout S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L., Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacle J.M., Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G., Reinert K., Remington K., Saunders R.D., Scheeler F., Shen H., Shue B.C., Siden-Kianci I., Simpson M., Skupski M.P., Smith T., Spier E., Spradling A.C., Stapleton M., Strong R., Sun E., Svirskaas R., Tector C., Turner R., Venter E., Wang A.H., Wang X., Wang Z.Y., Wassarman D.A., Weinstein G.M., Weissbach J., Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A., Ye J., Yeh R.F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L., Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Smith H.O., Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT "The genome sequence of *Drosophila melanogaster*.";
RL Science 287:2185-2195(2000).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=22426065; PubMed=12537568;
RA Celniker S.E., Wheeler D.A., Kronmiller B., Carlson J.W., Halpern A., Patel S., Adams M., Champe M., Dugan S.P., Frise E., Hodgson A., George R.A., Hoskins R.A., Laverty T., Muzny D.M., Nelson C.R., Pacle J.M., Park S., Pfeiffer B.D., Richards S., Sodergren E.J., Svirskaas R., Tabor P.E., Wan K., Stapleton M., Sutton G.G., Venter C., Weinstein G., Scherer S.E., Myers E.W., Gibbs R.A., Rubin G.M.;
RT "Finishing a whole-genome shotgun: release 3 of the *Drosophila melanogaster* euchromatic genome sequence.";
RL Genome Biol. 3:RESEARCH0079-RESEARCH0079(2002).
RN [3]
RP SEQUENCE FROM N.A.
RX MEDLINE=22426069; PubMed=12537572;
RA Misra S., Crosby M.A., Mungall C.J., Matthews B.B., Campbell K.S., Hradecky P., Huang Y., Kaminker J.S., Millburn G.H., Prochuk S.E., Smith C.D., Tupy J.L., Whitfield E.J., Bayraktaroglu L., Berman B.P., Bettencourt B.R., Celniker S.E., de Grey A.D., Drysdale R.A., Harris N.L., Richter J., Russo S., Schroeder A.J., Shu S.Q., Stapleton M., Yamada C., Ashburner M., Gelbart W.M., Rubin G.M., Lewis S.E.;
RT "Annotation of the *Drosophila melanogaster* euchromatic genome: a systematic review.";
RL Genome Biol. 3:RESEARCH0083-RESEARCH0083(2002).
RN [4]
RP SEQUENCE FROM N.A.
RX MEDLINE=22426070; PubMed=12537573;
RA Kaminker J.S., Bergman C.M., Kronmiller B., Carlson J., Svirskaas R., Patel S., Frise E., Wheeler D.A., Lewis S.E., Rubin G.M., Ashburner M., Celniker S.E.;
RT "The transposable elements of the *Drosophila melanogaster* euchromatin: a genomics perspective.";
RL Genome Biol. 3:RESEARCH0084-RESEARCH0084(2002).
RN [5]
RP SEQUENCE FROM N.A.
RA FlyBase;
RL Submitted (SEP-2002) to the EMBL/GenBank/DBSJ databases.
RN [6]
RP SEQUENCE FROM N.A.
RA FlyBase;
RL Submitted (MAR-2004) to the EMBL/GenBank/DBSJ databases.
DR EMBL; AE003626; AAS64673.1; -;
FT NON TER 1 1
SQ SEQUENCE 1761 AA; 197521 MW; 62A5E16E6241B3F4 CRC64;
Query Match 22.5%; Score 1327.5; DB 2; Length 1761;
Best Local Similarity 24.7%; Pred No. 1.1e-17;
Matches 451; Conservative 114; Mismatches 369; Indels 895; Gaps 85;

| | | | | |
|----|-----|--|---|------|
| QY | 74 | CF | -----ESFERGECDCDAOCKYDKCPDYESFC | 104 |
| DB | 5 | CFIHALYHSANAALGSEIKFGDESTESA | -----KY-----PDY-----CWIN | 45 |
| QY | 105 | | -----AEVNPTSPPS | 115 |
| DB | 46 | PFLPGCFPGDGNSTKISTLATTKSTVTSEITLKITIKSTAKETQKTNEPT | ----- | 105 |
| QY | 116 | SKAPPSSGA | -----SOTIKST--TKR | 135 |
| DB | 106 | TEKITTPKATIKSTATTARATAPKTEQTLTKTIKSTSELITLUKTIITIKSTAE | ----- | 165 |
| QY | 136 | SPKPPNKKTKVIESEIEITERKVDNKNRKKKTPKPPVVDVDEAGSLDNGDFKVT | ----- | 193 |
| DB | 166 | STHNPTTKSTLTTRTEETTRKSTAK--TTREPTTK | -----RETTTER | 207 |
| QY | 194 | -----PDTS--TTOHNVSTSKITITAKINPRPSLPNSDTSKETSULTVNKEITVE | ----- | 243 |
| DB | 208 | TTQEPSTSKTTTTH | -----TTABPATKKTTHTEPT--TOKSTTLRITTEETTRKSSSTA | 257 |
| QY | 244 | --TKETITTTINKQISTDCKEKTTSKAKTOSIEKTSKADLAPTSKVLAKPTKAKTITTKGPA | ----- | 301 |
| DB | 258 | KTIREPTTKRETTERTTKPST | -----KITTHE--TTAEPATKKTTHTEPTQKSTT | 307 |
| QY | 302 | LITPKPT | -----PTTPKEPASTTPKEPTPTTIKSAPTTPKEPA | 342 |
| DB | 308 | LRITEPTTRKSTARTTRETPTTKRETTERTTQEPSTSKTKTHETT--AEPATKKTTHTEPT | ----- | 366 |
| QY | 343 | TTKSA | -----PTTPKEPAPTTPKEPAPTTPKEPAPTTPKEPA | 379 |
| DB | 367 | TOKSTTLRITTEETTRKSTAKTIRE--PTTKRETTERTTKPSTKTTTHTAABPATK | ----- | 424 |
| QY | 380 | | -----PTTPKEPAPTTPKKA | 402 |
| DB | 425 | KTTHTEPTQKSTTLRITTEETTRKSTARTTRETPTTKRETTERTTQEPSTSKTKTHETTA | ----- | 484 |
| QY | 403 | | -----PTTPKEPTTPKEPAP | 436 |
| DB | 485 | EPATKKTTHTEPTQKSTTLRITTEETTRKSTAKTRETPTTKRETTERTTQEPSTSKTTT | ----- | 544 |
| QY | 437 | PKEPAPTAPKKA--PTTPK | -----EPA | 473 |
| DB | 545 | HETTAETATKKTTHTEPTQKSTTLRITTEETTRKSTAKTRETPTTKRETTERTTQEPST | ----- | 604 |
| QY | 474 | | -----PTTPKKA | 490 |
| DB | 605 | SKTTTHETTABPATKKTTHTEPTQKSTTLRITTEETTRKSTAKTRETPTTKRETTERTT | ----- | 664 |
| QY | 491 | | -----TTTKSA | 510 |
| DB | 665 | QEPSTSKTTTHETTABPATKKTTHTEPTQKSTTLRITTEETTRKSTAKTRETPTTKRET | ----- | 724 |
| QY | 511 | SPTTKKEPA | -----PTTPK | 529 |
| DB | 725 | TERTTKETPTTKTTTHKTTTEPTTKKTTHTEPTTKKSTTLAKTEPTTRKTSTTKTTEPT | ----- | 784 |
| QY | 530 | TPKKAAPTTPKEPA | -----PTTPKEPAPTTPKKA--PTAFK | 567 |
| DB | 785 | TKRKTERTKPTTKTTTHKTTTEPTTK | -----TTTKKTTHTEPTTKKSTTLKPTTEPT | 840 |
| QY | 568 | | -----PTTPKETA | 583 |
| DB | 841 | TRKSTTKTRETPTTKRETTERTTQEPSTSKTTTHETTABPATKKTTHTEPTQKSTTLRI | ----- | 900 |
| QY | 584 | ---PTT--PEKLAPTPEKPAPTTPEELAPTTPEPT | -----PTTPE---EPA-- | 623 |
| DB | 901 | TEBPTTRKSTAKTIRE--PTTKRETTERTTKETTRKTTTHKTTTEPTTKTTTHTEPTT | ----- | 957 |
| QY | 624 | | -----PTTPKAAAPTTPKEPAPTTPKEPAPTTPKKA | 660 |
| DB | 958 | KKSTTLKPTBPTTRKSTTKTIRE--PTTKXTTERTTKEPTTRKTTTHKTTTEPTTK | ----- | 1015 |
| QY | 661 | ETA | -----PTTLKEPAPTTPKAPKELA | 701 |

```

DB 1016 TTGKTTTHTTTKSTLTKPTEPTTRKSTTKTTRTEPTTKRVTTERTTREPSTTKTTH 1075
QY 702 TTSDKFA-----PTTPKGTA-----PTTKPEPAPTPKEPA-----PTTP 736
DB 1076 ETTAEPAKTKTHTTETQKSTLRLTEPTTRKSTAKTTEPTTKRETTERTKEPTR 1135
QY 737 KGTAPTTKPA-----PTTPK-----KPAPKELAPTTTKGPTSTTSDDKAPPTPKETA 785
DB 1136 KTTHTKTEPTTKTHTTETPKSTLTKPTEE-----PTTRK-----TSTTKTREPTEKKT 1190
QY 786 PTPKEPA-----PTTPK-----KP-----APTPEP 807
DB 1191 ERTTEPTTRKTTHTKTEPTTKTTHKTTHTTETPKSTLTKPTEPTTRKSTTKT 1250
QY 808 PPPTSEVSTPTTKP-----TTIKSPDE-----STPELSABPTPKALENSKEPGVP 857
DB 1251 REPTTKRVTTERTTREPTRKTTHTKTEPTTKTTHKTTTKTTHKTEPTTK-STILKPEEP 1309
QY 858 TTTPAATKPEMTTAKDKTTERDLR-----TTPK-----TTPK 887
DB 1310 TTKTSTTKTREPTRKRVTTERTTREPTRKTTHTKTTHTKTEPTTKTTHKTTHTK 1369
QY 888 TTTAAP-----KMTKETAT-----TTEKTESKITAITQVTSITQDTPPKIT 932
DB 1370 STILKPEPTTRKSTTKTTRTEPTTKRVTTERTTREPTRKTT--THKTEPTTKT 1427
QY 933 TLKT-----TTLAP-----KVTTK-----KTITTEIMNKP-----EE 961
DB 1428 TKKTHTTETPKSTLTKPTEPTTRKSTTKTTRTEPTTKRVTTERTTREPTRKTTHTK 1487
QY 962 TAKPKDRATNSKATPKPQ-----KP-----TKAPKKPT-----990
DB 1488 TREPPTTKTTHKTTETPKSTLTKPTEPTTRKSTTKATREPTRKRVTTERTTREP 1547
QY 991-----STKKPKMPRVKPKYTPPRKMTSTMPELNP 1022
DB 1548 TTRKTTTAAHTTEPTTKTTHKTTETPKSTLTKPTEPTTRKSTTKL-----KTEPTTRKSTTKTREP 1603
QY 1023 TSRIAEAMLOTT-----TRNQTPNSKLAVNPXSEDAAGAGET-----1062
DB 1604 TTRETSTVKTTADQTTKRTTAESMTNQETSVETTNSSNQNTTETSTTBEQVHH 1663
QY 1063-----PHMLLRPH-----1070
DB 1664 HHHHHYHKPADLGSILPLPLPLPLPLPLPLPLPLPLPLPLPLPLPLPLPLPLPL 1723
QY 1071-----VFMPREVTPDMVDYLPVNP 1088
DB 1724 PPLPEVNLTALSLPEISLNPPLPQLPN 1752

RESULT 10
Q7PMD5
ID Q7PMD5 PRELIMINARY; PRT; 3150 AA.
AC Q7PMD5,
DT 01-MAR-2004 (TremBLrel. 26, Created)
DT 01-MAR-2004 (TremBLrel. 26, Last sequence update)
DT 01-MAR-2004 (TremBLrel. 26, Last annotation update)
DE ENSANGP0000004655 (Fragment).
GN Name=ENSANGG0000003651;
OS Anopheles gambiae str. PEST.
OC Eukaryota; Metazoa; Arthropoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Nematocera; Culicoidae; Anopheles.
OX NCBI_TaxID=180454;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=PEST;
RA Anopheles Genome Sequencing Consortium;
RL Submitted (APR-2003) to the EMBL/GenBank/DBJ databases.
CC !- CAUTION: The sequence shown here is derived from an
CC EMBL/GenBank/DBJ whole genome shotgun (WGS) entry which is
CC preliminary data.

```

```

DR EMBL; AAAB0100980; EAA13969.2; --
DR GO; GO:0016020; C:membrane; IEA.
DR GO; GO:0004872; F:receptor activity; IEA.
DR GO; GO:0008151; P:cell growth and/or maintenance; IEA.
DR InterPro; IPR006770; OGF:recept.
DR Pfam; PF04680; OGF_I_III; 80.
FT NON_TER 1
FT NON_TER 3150 3150
SQ SEQUENCE 3150 AA; 322879 MW; 3C7B3D441CE8C839 CRC64;

Query Match 21.5%; Score 1266.3; DB 2; Length 3150;
Best Local Similarity 20.1%; Pred. No. 3.2e-16;
Matches 411; Conservative 110; Mismatches 416; Indels 1107; Gaps 76;

QY 110 PTPSP-----SSKAPPPSGASQIKSTKSP-KPPNKKTKKVISEETEKVK 159
DB 832 PTPSTSTSTDTMTSSASPEPS-----TTPGTRTTRPTSTSTSTDTMTSSASPEST 887
QY 160 DNKKNR-TKKKTPKPPVVDKAGS-----GLDNGDFKVTTPDTS 197
DB 888 TSGTTRTTRPTDITMTSSASPESTTPGTRTTRPTSTSTDTMTSSASPEPS 947
QY 198 TTQHN-----KVSTSKITAKPINP-----218
DB 948 TTPGTRTTRPTSTSTDTMTSSASPESTTPGTRTTRPTSTSTDTMTSSAS 1007
QY 219-----RPSLPPNSDTSKTSLVKNKETTVEKETTNNKQSTDKGKTT 263
DB 1008 PPSKPGTTRTTRPTSTSTDTMTSSASPESTTPGTRTTRPTSTSTDTMT 1066
QY 264 SA-----KETQSIKTSKADLAP-----TSKVAKPTPKAET 295
DB 1067 SGASTPESTTPGTRTTRPTSTSTDTMTSSASPESTTPGTRTTRPTSTST 1126
QY 296-----TTKGPALTTPKEPTP-----TTPKEPAS 318
DB 1127 DTTMSSASPESTTPGTRTTRPTSTSTDTMTSSASPESTTPGTRTTRPT 1186
QY 319-----TTPKEP-----TPTTKSAPT 334
DB 1187 TSTSTDTMTSSASPESTTPGTRTTRPTSTSTDTMTSSASPESTTPGTRT 1246
QY 335 TP-----KEPA--PTTKSAPT-----351
DB 1247 RPTSTDTMTSSASPESTTPGTRTTRPTSTSTDTMTSSASPESTTPGTRT 1306
QY 352-----KEPAPTTPKEPAPT-----TTPKEPA--PTTKSAPT 383
DB 1307 TRPTSTSTDTMTSSASPESTTPGTRTTRPTSTSTDTMTSSASPESTTP 1366
QY 384 KSAPTTP-----KEPAP-----TTPKAPAPT-----TTPKEPAPTTP 414
DB 1367 RTTPRPTSTSTDTMTSSASPESTTPGTRTTRPTSTSTDTMTSSASPE 1426
QY 415 KEPTTPTPKEPAPT-----TTPKEPAP-----TTPKEPAPT-----APKKA 449
DB 1427 PGTRTTRPTSTSTDTMTSSASPESTTPGTRTTRPTSTSTDTMTSSASPE 1486
QY 450 P-----TTPKEPAP-----TTPKEPAPTTPKEPAPTTP 477
DB 1487 PSTTPGTRTTRPTSTSTDTMTSSASPESTTPGTRTTRPTSTSTDTMT 1546
QY 478 KEPA-----PTTKSAPT-----TKEPA--PTTKSAPT-----507
DB 1547 -PESMTPGTRTTRPTSTSTDTMTSSASPESTTPGTRTTRPTSTSTDT 1605
QY 508 --KEPSTTPKEPAPTTPKEPAP-----TTPKAPAPT-----TTPKKA 536
DB 1606 SASTPESTTPGTRTTRPTSTSTDTMTSSASPESTTPGTRTTRPTST 1665
QY 537 -----TTPKEPAPT-----TTPKEPAPTTPKKA-----PTA 562
DB 1666 SASTPESTTPGTRTTRPTSTSTDTMTSSASPESTTPGTRTTRPTST 1725

```



```

QY 129 -----IKSTTKRSPKPNKKTKKVIIESEBITKVKDNKNRTKKAPT 171
Db 544 GICDPSNPFRPPQPPSPPPPPRP-----PRAPR 577
QY 172 PKPPVVDAGSLDNGDFKVTPTD-----STQHNVKSTSPKLTITAKPINPRPSLPNSDT 228
Db 578 PSPP-----FHPSPDPASVPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSP 620
QY 229 SKETSUTVNKETTIVETKTTTNTKOTSDGKEKTTSAKETOSIEKTSKDLAPTSKVLA 288
Db 621 -----PSPSPAPPS 628
QY 289 PTPKABTTTKGALITTPKEPTTTPKEPAST-----TPKEPTTTPKSAPTTPKEPAP 341
Db 629 PAPP-----PAPPSPQPSVPQPSVPQPSVPQPSVPQPSVPQPSVPQPSVPQPSVP 682
QY 342 TTTKGAPTTPKEPATTTPKEPATTTPK-----EPAPTTPKEPATTTPKSAPTTPKEPAPT 397
Db 683 APPNPAPPSPAPPLPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSP 742
QY 398 PKKPAPT-----TPKEPAPTTPKEPTTTPKEPAPTTPKEPAPTTPKEPAPT 443
Db 743 PEPSPAPPSPPPSPAPPSPAPPSPAPPSPAPPSPAPPSPAPPSPAPPSPAPPSPAPP 802
QY 444 ----AFKKPAPTTPKEPAPTTPKEPAPT-----TTK 470
Db 803 PPSAPPSPSPSPAPPSPSPSPAPPSPSPAPPSPSPSPAPPSPSPSPAPPSPSPAPP 862
QY 471 EPSPTTPKEPATTTPKSAPTTPKEPATTTPKSAPTTPKEPATTTPKEPATTTPKEPAPT 528
Db 863 SPAPPSPSPSPSPAPPSPSPSPAPPSPSPSPAPPSPSPSPAPPSPSPSPAPPSPSP 921
QY 529 -----TTPKPAPTTPKEPAPTTPKEPAPTTPKKAAPTAPK-----564
Db 922 PSPEPSPSPSPAPPSPSPSPAPPSPSPSPAPPSPSPSPAPPSPSPSPAPPSPSPSP 981
QY 565 -----EPAPTTPKETAPTTPKKTTPPTPEKLAPTTFE--KPAPTTPKEPAPT 609
Db 982 DPSPAPPSPDPSPAPPSPAPPSPAPPSPAPPSPAPPSPAPPSPAPPSPAPPSPAPP 1040
QY 610 -----TBEPTPTTPEEPAPTTPKKAAPN-----TPKEPAPTTPKSPAP 648
Db 1041 SPSPSPSPSPSPAPPSPAPPSPAPPSPAPPSPAPPSPAPPSPAPPSPAPPSPAPP 1100
QY 649 TTPK-----EPAPTTPKETAPTTPKGTAPTL-----TKPTSTTS 704
Db 1101 APPSPSPSPAPPSPAPPSPAPPSPAPPSPAPPSPAPPSPAPPSPAPPSPAPPSP 1160
QY 676 -----KEPAPTTPKKAPELAPT-----TKPTSTTS 704
Db 1161 VPPTSPSPAPPSPAPPSPAPPSPAPPSPAPPSPAPPSPAPPSPAPPSPAPPSPAPP 1220
QY 705 DKPAPTTPKGTAPTTTPKEPAPTTPKGTAPTTTPKEPAPTTPKKAPELAPT--763
Db 1221 PSPAPPSPAPPSP-APPSPAPPSPAPPSPAPPSPAPPSPAPPSPAPPSPAPPSPAPP 1275
QY 764 TTKGTSTTSKPAPTTPKETAPTTPKEPAPTTPKKAPELAPTTPPTT-TSEVSTPTTK 822
Db 1276 PSPAPPSPAPPSPAPPSP-APPSPAPPSPAPPSPAPPSPAPPSPAPPSPAPPSPAPP 1332
QY 823 EPTTHKSPDESTPLSAEPTPKALE-----NSPKPCGV---PT 858
Db 1333 PPLPSPAPPLEVPASAPSPSPPLRPQPTAPMPSPAPPSPAPPSPAPPSPAPPSPAPP 1392
QY 859 TKTP-----AATKPE-----868
Db 1393 TPTPLAPLPDCTLLAQAALLSPDAANSFVVSAGLFTSVAPSTTPPELLASCTVCSC 1452
QY 869 -MTTFA-----873
Db 1453 QLTATAISLVGSSKGNNTNGSSGNGNYNGGDAATQIRPAGNSTDRGSSSSGSSSS 1512

```

```

QY 874 -----KDKTTERDL-----RTTP-----886
Db 1513 WGPGETATAEAWDAVDGTYQLQSLIGGVYTRTVVVDRTTPSVSGNVTLSANRIKQEP 1572
QY 887 -----ETTTAAPKVT-----896
Db 1573 SAVGEASLNALGSKQAMLLTISFSEVPAPDPAASLIIVTALVAEWAADKMTFVFLAM 1632
QY 897 ---KETATTTEKTESKIT-----912
Db 1633 TLPABELVATAAGSSSSGTSRSGNGNGTAAAAAAPPAGTTGRRALOOQAAAPPP 1692
QY 913 -----912
Db 1693 ASGSSSLSGAATANQQORHVHFLPATAYADAARNPGRNDLSLSVELTDNAVAPVCE 1752
QY 913 ---ATTQVSTTTQDTPPKITTLKTTTLAPKV-----944
Db 1753 ALATTARVTAATYP-----AVAATTTLVAAAASSSFAQAIRAKGSLQGSYHIQMLTMS 1805
QY 945 -----TTKKTITTE-----954
Db 1806 LYLASKGVREGEYAVEFKYAVLVGKGNLGPAAEMPTNEKEVTAAEQARQVGGDLWPI 1865
QY 955 ---IMNKPETETAKPRATNSKATTPKQKP-----982
Db 1866 GNDLLGSSNTTASGSSSGSSSSNSP-PRRPPPPPAAGSTGLLFSNADASPPPLAVATP 1924
QY 983 -----TPAKPKPTSTKKPTMPRVKPTTTPRKMTST--MBELN-----1021
Db 1925 ALPAPLESTIAAATAAP-----PKLPSP---PPPAVGSSGTVLPRRHLMOOM 1969
QY 1022 -----PTSRIAEAMLQTTTRPNQTPNSKLVE-----1047
Db 1970 LQPPAAVAAPPPPPASSALVLQSPPPPPPPSQLLIQQASATVSDMQDLYTLVAA 2029
QY 1048 -----VNPKESEDAGAGSETFHMLLRPHVPMVEVTPDMVLP--1085
Db 2030 MLITAVAAGRLIAAVLYRLVSP-----PHPFLA-----FPXLE 2064
QY 1086 -----VP 1087
Db 2065 TITAGLILVALTFYSMALGGPAADWHGSRTAAYCVLITAVVYAAFLWLALARAWVP 2124
QY 1088 NQGIINPM 1096
Db 2125 -QFTLVEPM 2132

RESULT 12
AAS07044
ID AAS07044 PRELIMINARY; PRT; 3409 AA.
AC AAS07044;
DT 02-MAR-2004 (TrEMBLrel. 27, Created)
DT 02-MAR-2004 (TrEMBLrel. 27, Last sequence update)
DT 02-MAR-2004 (TrEMBLrel. 27, Last annotation update)
DE plus agglutinin.
GN SAGI.
OS Chlamydomonas reinhardtii.
OC Eukaryota; Viridiplantae; Chlorophyta; Chlorophyceae; Volvocales;
CC Chlamydomonadaceae; Chlamydomonas.
CX NCBI_TaxID=3055;
RN [1]
RP SEQUENCE FROM N.A.
RA Ferris P.J., Waffenschmidt S., Umen J.G., Ishida K., Kubo T., Iau J.,
RA Goodenough U.W.;
RT "Plus and Minus Sexual Agglutinins from Chlamydomonas reinhardtii.";
RL Submitted (OCT-2003) to the EMBL/GenBank/DBJ databases.
DR EMBL; AY450930; AAS07044.1;
SQ SEQUENCE 3409 AA; 336045 MW; 7FE87633EDD6631F CRC64;

Query Match 20.5%; Score 1205.1; DB 2; Length 3409;
Best Local Similarity 15.2%; Pred. No. 5.2e-15;

```


QY 1048 -----VNPKSEDAGAGETPHMLLRPHVFMVPTPMDYLP-- 1085
 Db 2030 MLITAVAGRLTAANVYRLVSP-----PFPFLA-----FPRLE 2064
 QY 1086 -----VP 1087
 Db 2065 TTIAGLIVALTIFYSCMALGSPADWHGSRTAAYCVLTIAVYVYAAFLWLALARAMVVP 2124
 QY 1088 NGGIINPM 1096
 Db 2125 -QFTLVEPM 2132

RESULT 13
 ID SLIP1_CLOTM STANDARD; PRT; 1664 AA.
 AC Q06852; 1994 (Rel. 29, Created)
 DT 01-JUN-1994 (Rel. 29, Last sequence update)
 DT 05-JUL-2004 (Rel. 44, Last annotation update)
 DE Cell surface glycoprotein 1 precursor (Outer layer protein B) (S-layer protein 1).
 DE Name=olpB;
 GN Clostridium thermocellum.
 OS Clostridium thermocellum.
 OC Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
 OC Clostridium.
 OX NCBI_TaxID=1515;
 RN [1]
 RC SEQUENCE FROM N.A.
 RC STRAIN=NCIB 10682;
 RX MEDLINE=93209931; PubMed=8458832;
 RA Fujino T., Beguin P., Aubert J.-P.;
 RT "Organization of a Clostridium thermocellum gene cluster encoding the
 RT cellulosomal scaffolding protein C1pA and a protein possibly involved
 RT in attachment of the cellulosome to the cell surface";
 RL J. Bacteriol. 175:1891-1899 (1993).
 CC -!- SUBUNIT: Assembled into mono-layered crystalline arrays.
 CC -!- SUBCELLULAR LOCATION: Cell wall.
 CC -!- SIMILARITY: Contains 4 S-layer homology (SLH) domains.
 CC
 CC This SWISS-PROT entry is copyrighted. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL; X67506; CAA47841.1; -.
 DR PIR; T18262; T18262.
 DR InterPro; IPR008965; Cellul_bind.
 DR InterPro; IPR001119; SLH.
 DR Pfam; PF00395; SLH; 3.
 DR PROSITE; PS01072; SLH DOMAIN; 2.
 DR Cell wall; Repeat; S-layer; Signal.
 KW SIGNAL 1 28
 FT CHAIN 29 1664
 FT DOMAIN 36 763
 FT REPEAT 36 191
 FT REPEAT 207 363
 FT REPEAT 409 565
 FT REPEAT 607 763
 FT DOMAIN 771 1377
 FT DOMAIN 1378 1449
 FT DOMAIN 1453 1494
 FT DOMAIN 1495 1565
 FT DOMAIN 1566 1625
 FT DOMAIN 1626 1646
 FT SEQUENCE 1664 AA; 178194 MW; 5F396695BA9FE74B CRC64;
 SQ

Query Match
 Best Local Similarity 19.2%; Score 1132; DB 1; Length 1664;
 Pred. No. 5.3e-14; Pred. 1664;

Matches 371; Conservative 124; Mismatches 421; Indels 820; Gaps 69;
 QY 4 KTIPIYLLILLVVFVQVSSQDLSSCAGCGEGYSDATCNCYNCQHYMECCP----- 58
 Db 6 KVLISILUTLL-----IISTTSVNM-----FAEATPSIEW 37
 QY 59 -----DFKEVCTAELSKRCFCSEFSGRECDCAQCKYDKCCPD 99
 Db 38 LDKTEVHVGDIATIKVNNIRKLAGYQLNIK----- 69
 QY 100 YESFCAEVHNPTSP-----PSSKAPPPSGASQTIKS----- 131
 Db 70 ---FDEVLQFVDPATGEEFTDKSMFVNRLTNSXYGTPV--AGNDIKSGINPATGYN 125
 QY 132 --TKGRSP-----KPPNKKTK----- 146
 Db 126 NLTAAYSSGIDEHTGIIIGIFKVLKKQNTSRFEDTSLMPGAISGTSLPDWAETITGY 185
 QY 147 -----KVIESEITE-----KVK-----DNKK-----RTKKKPT- 171
 Db 186 EVIQPOLIVVEABPLKDAVLELDKTKVKVGDIITATIKENMKNFAGYQLNIKVDPTM 245
 QY 172 -----PKPPVDEAGSGLDN----- 186
 Db 246 LEAIELETGSAIAKRTWPTVGTGTVLOSNDNYGKTTAVANDVGAGIINFABAYNLTKYRT 305
 QY 187 -----GDFKVITPDT-----STTQHNKV 204
 Db 306 GVAEETGIIGIKIGFVLKAGSTAIRFEDTTAMPGAIEGTWFDWYGENIKGYVWQPGI 365
 QY 205 -----SISPKITAKPNRPSLP-----PNS-----DTSK----- 230
 Db 366 VAEGEPGPEETPEPVTETPDVPTVTEEPVPSLPDSYVIMELDKTKVKVGIITAT 425
 QY 231 -----ETSLTVKETTETKETTNNKQSTQCKEKT 262
 Db 426 IKIENMKNFAGYQLNIKVDPTMLEAIELETGSAIAKRTWPTVGTGTV-----LQSDNYGKT 480
 QY 263 TS-----AKTSIETKSADKLAPTSKVLAKPTPKAET 295
 Db 481 TAVANDVGAGIINFABAYNLTKYRTGVAETGTGIGIKIGFVLKAGSTAI-----RPF 535
 QY 296 TT-----KGPALTTP-----KEPTPT-TPKEPASTPKEP 324
 Db 536 TTAMPGAIEGTWFDWYGENIKGYVWQPGIEVAEGEETPEPVTETPDVPTVTEEP 595
 QY 325 TPT----- 327
 Db 596 VPSELPDSYVIMELDKTKVKEGDIITATIRVANNIKLAGYQIGIKYDPKVEAFNIETGD 655
 QY 328 ----- 327
 Db 656 PIDEGTWPAVGGTILKNRDYLTGVAINNVSKILNFAAYVYVDDYEEGSEDTGIIG 715
 QY 328 -----TIKSAPTPK-----EPATPTTKSAPTPK 352
 Db 716 NIGFVLKAEADITIRFEELESMEGSDGTGYMLDWNLNRSYGVVWQPAIKAAS-----D 770
 QY 353 EPATPTTKEPATTPKEPAP-----TTKEPAPTTKSAPTTKEPATTPKAPATTPKE 408
 Db 771 EPIPTDTPSDEPTPSDEPTPSDEPTPSDEPTPSDEPTPSDEPTPSDEPTPSDEPTPSDE 830
 QY 409 PAPT-----TPK-ETPT-TPKEPAPT-TKEPAPTTKEPAPTPAKKPA--PTTPK 454
 Db 831 PPSDEPTPSDEPTPSDEPTPSDEPTPSDEPTPSDEPTPSDEPTPSDEPTPSDEPTPSD 890
 QY 455 EPAPT-TPKEPAPTTKPEPSTTPKEPAPTTKSAPTTKEPATTPKAPATTPKPEPST 513
 Db 891 EPTPSTPEPIPTDTPSDEPTPSDEPTPSDEPTPSDEPTPSDEPTPSDEPTPSDEPTPS 947
 QY 514 TKEPAPTTKEPAPTPKKA-----PTTPKEPAPT-TPKEPAPTTKPE 558
 Db 948 DTPSDEPTPSDEPTPSDEPTPSDEPTPSDEPTPSDEPTPSDEPTPSDEPTPSDEPTPSD 1007

| | | | | |
|----|------|---|----------------|------|
| QY | 559 | APTAPKEPAPTTPKERTAPTKKLTPTPPKLAPTTPKEKPAPT--- | TPEELAPTTPEEPT | 615 |
| Dd | 1008 | EPTPSDEFTESD-----EFTPSDETPSDETPSD-EPTSPDEPT | : : : : : | 1046 |
| QY | 616 | PT-TPEEPAPTTPKAANPTKPEAPTTPKEPA--PTTPKEPAPT-TPKTAPTTPKGTA | : | 671 |
| Dd | 1047 | PSETPEEPIPTDPSEDPTSDSEPTSDSEPTSDSEPTSDSEPTSEPIPTDP | SDE | 1106 |
| QY | 672 | PTTLKEPAPTTPKKAPKELAPITTKTEPT-STTSDKPAPTTPKGATATTPKEPAPTTPKE | | 730 |
| Dd | 1107 | PTPSDEFTPS--DEFTESD-EFTPSDEFTSETEEBIPTDPSEDPTSDSEPTPS--DE | | 1161 |
| QY | 731 | PAPTTPKGTAPTTLKEPAPT-TPKKAPAKELAPITTKGPTSTISDKCAPT--TPKETAP | | 786 |
| Dd | 1162 | PTPS----DEFTPSDEFTSETEEBI-----PDTPSDSEPTSDSEPTSDSEPTSD-EP | | 1211 |
| QY | 787 | TTPEKPAPT-TPKKAPATTETPPPTTVSEVSTPTTKKEPTTIHKSPDESTPELSABETPK | | 845 |
| Dd | 1212 | TPSDEPTSETEEBEPIPTDPSPDEFTPSD--EFTPSDEPT---PSDEPTP--SDEPTP- | | 1262 |
| QY | 846 | ALENSKPEPGVTTKTPAATKPEMTTAKOKTTERDLRTTETTAAAPKMKTETATTTEK | | 905 |
| Dd | 1263 | --SETPEP-IPTDPSEPTSDSEPTSDSEPTSDSEPTSDSEPTSDSEPTSDSEPTSPSE | | 1319 |
| QY | 906 | TTSEKITATTQTVSITTQDTTPPKITLTKTLLAPKVTTTKKITTITEIMNKPEETAKP | | 965 |
| Dd | 1320 | TPEEPIPTDPSEDPTSDSEPTSDSEPT-----PTTPTPTPTPTPTSGGGSG | | 1354 |
| QY | 966 | KDRATNSKATTPKQP-KPTKAPKKPTSTKKPKTMPRVKPKTTPTRKMTS---- | | 1015 |
| Dd | 1355 | SDREPTSDSEPTSDSEPTSPSTPEEPT-----PTTPTPTPTPTPTSGGGSG | | 1401 |
| QY | 1016 | -----TMPELNPTSRIAEAMILQTTTRPNCTPNSKLVEYNPKSEDAGGAEGE | | 1061 |
| Dd | 1402 | GSGGGGGGGGTVPSTPTPTS-----KPTSTPAPTIE-EPTPSOVPGAIGG | | 1449 |
| QY | 1062 | TPHMLLRPH--VFMPH----- | | 1075 |
| Dd | 1450 | EHRAYLRGYPDGSPRPERNITRAEAIVIFAKLLGADSESYGAQSASPYSDLADTHWAAWAI | | 1509 |
| QY | 1076 | -----VTPD-----MDYLFRVNPQGI-----IINP | | 1095 |
| Dd | 1510 | KFATSQGLFKGYPDGTFKPDQNTIRAEFATVWLHLFTKVKQEIMSKLAITIDISNP | | 1565 |

RESULT 14

| | | |
|--|--------------|---------------|
| Q6SSE8 | PRELIMINARY; | PRT; 3889 AA. |
| ID AC Q6SSE8; | | |
| DT 05-JUL-2004 (TrEMBLrel. 27, Created) | | |
| DT 05-JUL-2004 (TrEMBLrel. 27, Last sequence update) | | |
| DT 05-JUL-2004 (TrEMBLrel. 27, Last annotation update) | | |
| DE Minus agglutinin. | | |
| GN Name=SADI; | | |
| OS Chlamydomonas reinhardtii. | | |
| OC Eukaryota; Viridiplantae; Chlorophyta; Chlorophyceae; Volvocales; | | |
| OC Chlamydomonadaceae; Chlamydomonas. | | |
| RN NCBI_TaxID=3055; | | |
| RX [1] | | |
| RP SEQUENCE FROM N.A. | | |
| RC STRAIN=CC-621; | | |
| RA Ferris P.J., Waffenschmidt S., Umen J.G., Ishida K., Kubo T., Lau J., | | |
| RA Goodenough U.W.; | | |
| RL Submitted (OCT-2003) to the EMBL/GenBank/DDBJ databases. | | |
| DR EMBL; AY450929; AASG7042.1; -- | | |
| DR Interpro; IPR006315; Autotransporter. | | |
| DR Interpro; IPR003985; CoxA-like lec gl. | | |
| DR Interpro; IPR003882; Pistil_extensin. | | |
| DR PRINTS; PR01218; PSTLEXIENSIN. | | |
| DR TIGRFAMs; TIGR01414; autotrans_barl; 1. | | |
| SQ SEQUENCE 3889 AA; 389223 MW; 4DB5B44D5507214A CRC64; | | |


```

QY 66 -----AEISC-----KGRCF-----75
Db 252 PEFSLVSELACLFLEIFLEDVALRADYSVLMAQNTTLDNGKVFNLDPSTFNVNVL 311
QY 76 -----ESPERG-----REC-----D 85
Db 312 YNNVFMLEIPSPMLHSSVSVKASSFNFGALRNLARTCNFAPFTELHAMRWGQAQUD 371
QY 86 CDAQ-----89
Db 372 WSAEYQAALSYVDGSQLRGRWLPPVIBPVNSARLLLPTLRLRLTIIRTRDQHGAVQWT 431
QY 90 -----CK 91
Db 432 SRPLITGLPGEWALLRNLEYLDLSDMETGAIVGPISTWLMWHLRVINMTGHNFCKR 491
QY 92 KYDK-----95
Db 492 DWHKILISQIRMYRAATHEPNLNVPHYYGPGNGNNGMTRYNISVYDLSHGHWQWYDEV 551
QY 96 -----CC-----PD-----Y--ESFC 104
Db 552 TTEAGFEVIAPHGQCCKWKSQTIKDNVVEILYPGSRFGNNVQDEIYGGFYQDEWC 611
QY 105 AEVHNPTS-----PSSKKAP-----120
Db 612 ----EFTSPQPPPPAPSPSPPTTDPVPMPPSSPPAPVMPAPPQPPIPPASPLT 667
QY 121 -----120
Db 668 PAAAPRPLPTWPGKWEAGWPRPIPPRPRPPPPPLPPSPPLPPVPTSPSPPPPPPK 727
QY 121 -----120
Db 728 SPPPPKSPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPSPG 787
QY 121 -----PSSGA--SOTIKSTTKRSPKPNKKTKV-----146
Db 788 PPSPEPPSPAPSAAPSPSPSPAPSPDPSPKPPSPVPPSPPLPPSPSPSPVPPSP 847
QY 149 ISEEEITEKVKONKQRTKKKTPKPPVVDEAGSLDNGDKVTTPTDSTTTOHNVSTSP 208
Db 848 PASPEPTSPAPSPSPSPSPSPAPSP-----SPPPSPSPSPSPAPSP 891
QY 209 KITTAKPINRPSLPSNDSKETSITVNKETTVEKETTITNKQTSIDGKETSASET 268
Db 892 PPSPEPPSPAPPLPP-----907
QY 269 QSIKETSADLAPTSKVLAKTPKAEITTKGPALTTPKEPTTTPKEPASTTPKEPTPT 328
Db 908 ----PPSPPPSPAPSPPPSPSPAPSPPPSPSPAPSPPPSPAPSPPP-PPSPSPAP 962
QY 329 IKSAPTTKEPAPTTKSAPTTKKEPAPTTKEPAPTTKEPAPTTKEPAPTTKSAPT 388
Db 963 PLPPPPSPSPAPSPPPSPSPAPSPAPLPPSPSPSPAPSPPPSPSPAPSPPPSPAP 1022
QY 389 TKKEPAPTTKKEPAPTTKEPAP-----TTPKEPTTTPKEPAPTTKEPAPTTKEPAPTA 444
Db 1023 SPSPSP-EPSPAPSPPPSPSPAPSPAPSPAPSPAPSPAPSPAPSPAPSP-A 1080
QY 445 PKKAPATTK--EPATTKEPAP-----TTTKEPPTTPKEPAPTTTTSAPTTTKEPAPT 498
Db 1081 PPSAPSPSPAPSPAPSPAPSPAPSPAPSPAPSPAPSPAPSPAPSPAPSPAPSPAP 1140
QY 499 TTKSAPTTKEPSPITTKKEPAPTTKEPAPTTKKEPAPTT-----TPKEPAPTTKEPAP 552
Db 1141 PPSPPPPSPSPAPSP-----PPSPSPSPSPSPSPSPSPSPSPSPSPAPSPAPSPAPLPLPSP 1198
QY 553 TTTKKEPAPTAPKEPAPTTPKETAPTTPKLTTPTEKLAPTTPKEPAPTTPBELAPTPE 612
Db 1199 HTQSPSP-VPSPAPSPAPSPSPAPSPAPSPAPSPAPSPAPSPAPSPAPSPAPSP 1255

```

```

QY 613 EPTTTPPEEPAPTTPK-----AAANPTKPEPAPTTPKPEPAPTTPKETAAPTTPK 668
Db 1256 PFPSPSAPPSPPTTPSPPEPAPQPPSTPHAPPSPPEPSP-TPSPPLPPSPPEPSPPEP- 1313
QY 669 GTAPTTLKEPAPTTPKPAKELAPTTTKETSTTSKPAPTTPKGTAPTTKKEPAPTTP 728
Db 1314 -----SPAPSPSPSP-----APPSPMPPS-----PAPLAPQPPSP-TPSPSPPEPVP 1355
QY 729 KEPAPTTPKGTAPTTLKEPAPTTPKPAKELAPTTTKGPTSTTSKPAPTTPKETAAPT 788
Db 1356 PSPPEPVPDPDPPLPSPPTPPSPQPPVPSPPTPPSPQPPS-----PAPSP-----APSA 1406
QY 789 PKEPAPTTPK-----KPAPTTPETPPPTSEVSTTTTKETPTTHKSPDESTPELSABPTP 844
Db 1407 PLOPSPPPSPQPPSPAPGPPSP-----SPSTSPSPAPLAPAPVPVPMAPQPPS 1460
QY 845 KALENSPKPEGPVTTKPAATKPE-----868
Db 1461 PPLSPPPPPPPSPPTIIPASQPAPAAAALDCSAAATRTSFVASSSRGAFYIAVAPAS 1520
QY 869 -----868
Db 1521 SPSPYQVGCCELSYAVLDPGASQOYVIPSSGSSSTAGGSPTVAVTSSVTPAGAGGLNG 1580
QY 869 -----MTTAKDKTTERDLR-----883
Db 1591 THGSTARRRALLVEATASSGPAAGVGARHLLLAATANSTTLEGLATGRSRSAAGAGMG 1640
QY 884 -----TTPETT-----889
Db 1641 MSRLQVVDVQGLDPPVTAAPPTGTSNTTSGAGEAGSGTVRYSSMGAGSGGLDAWR 1700
QY 890 -----889
Db 1701 LTPGATGYLLRLKVADQEWWRWYSDIDPPRAAGQLLLARRTGGSSSSNSTSGSALA 1760
QY 890 -----889
Db 1761 AADEVQVENVNAAAGATAVSAQAQAAVRLMAMVAMSEFVQPFSTSLRLSGGARLL 1820
QY 890 -----TAAPKM-----TKETATTEKTESKITATTQVTSITT-- 923
Db 1821 STQCFASATAAAEVAAGAVTADSTPPGSDASSAATVAPAAIAPVSGTSGTSPNSSTSG 1880
QY 924 -----QDTTPKTIITLKTIT-----LAPKVTITKTIITTIMN 957
Db 1881 SAYQSCVAVLFAEQDATPELLUPPGTLTDMHGNINAEPLILSVNLTASADSUSTVERAG 1940
QY 958 KPEETA-----KPKOR 968
Db 1941 APVAAAAGGVFAASAFTSASASPLSAFSSRSLLOSGYHIQMLAMSSSLASPGISAFR 2000
QY 969 -----ATNSKA-----974
Db 2001 RISYLRWSLIGIOGNIPLLDGFSGSAAAGSGSGSSSSSSGGLGDVDAVALDRL 2060
QY 975 --TTPK-----QKP-----982
Db 2061 QLSVPPPLPAAGDAASQAQPPANLSPPSASOLVADGSTALAGRSRSLVQAARVAPSP 2120
QY 983 --TKAPKKP-----TSTKKPKTMPR-VRKPKTTP-----1008
Db 2121 PFTQAPTTPAPFGTGAAPAPPPAPPPQSPSTPPPPAPPPMPALSLSGDRDVLVSMLOQVGA 2180
QY 1009 -----TPRKMTSTPE-----LNP-----1022
Db 2181 IGSNSNGSGGSATSDASVAYSLOGAAASPRGDVVVLPDGGQLGLLGPGGSPPPMQPG 2240
QY 1023 -----TSRIAEAMLQTTTRPN-----1038
Db 2241 SGSGGQSGSGASVATDTIHTNVQDILTYLAIALLMVALVAAHLLVIGLRLVAMVYDVC 2300
QY 1039 -----QTENSKLV-----1046

```

```
Db      2301 GAAESGVEGLHPEVLRPPRAEMVILGGLLVALTFYSALTLSGAASPRWGDNTAAGRLIAVL 2360
Qy      1047 ----- 1046
Db      2361 VLAVLVVPYGLLLWMLTVCRWYLQEEVDHYMLGPHWQAFDGVIPGGAGAGDGGGGHGA 2420
Qy      1047 -----EVPKSEDAAGGAET-----PHMLLRP 1069
Db      2421 SALPAVAGFGTGGGGVAFACQAEPSGGGGGGEDGYGLGPHWALAP 2471
```

Search completed: October 13, 2004, 12:17:37
Job time : 156.725 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: October 13, 2004, 11:23:49 ; Search time 22.3651 Seconds
(without alignments)
4723.689 Million cell updates/sec

Title: SEQ1-C
Perfect score: 5887
Sequence: 1 MAWKLPYVLLLSVFVIQ.....DMDYLPVNPQGIINPMLS 1098
Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.1

Searched: 283416 seqs, 96216763 residues

Total number of hits satisfying chosen parameters: 283416

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR 79: *
1: pir1: *
2: pir2: *
3: pir3: *
4: pir4: *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Match | Length | DB ID | Description |
|------------|--------|-------|--------|----------|--------------------|
| 1 | 1325 | 22.5 | 3020 | 2 A43932 | mucin 2 precursor, |
| 2 | 1132 | 19.2 | 1664 | 2 T18262 | S-layer protein - |
| 3 | 1020.9 | 17.3 | 1889 | 2 T31108 | cyst germination s |
| 4 | 1010.7 | 17.2 | 7862 | 2 I38346 | elastic titin - hu |
| 5 | 1008.3 | 17.1 | 1274 | 2 T16251 | hypothetical prote |
| 6 | 996.8 | 16.9 | 2187 | 2 T30826 | nascent polypeptid |
| 7 | 992.1 | 16.9 | 3570 | 2 T45025 | mucin MUC5B, trach |
| 8 | 923.9 | 15.7 | 1367 | 1 S48478 | glucan 1,4-alpha-g |
| 9 | 910.5 | 15.5 | 3507 | 2 T34513 | hypothetical prote |
| 10 | 903.1 | 15.3 | 1188 | 2 S49915 | extensin-like prot |
| 11 | 872.9 | 14.8 | 6842 | 2 T29757 | protein UNC-89 - C |
| 12 | 871.5 | 14.8 | 1229 | 2 T25697 | hypothetical prote |
| 13 | 861.9 | 14.6 | 1151 | 2 T18535 | high molecular mas |
| 14 | 858.5 | 14.6 | 2897 | 2 B48666 | cell proliferation |
| 15 | 855.1 | 14.5 | 3256 | 2 A48666 | cell proliferation |
| 16 | 840.3 | 14.3 | 4135 | 2 T42629 | tenascin-X - bovin |
| 17 | 835.2 | 14.2 | 5762 | 2 A41819 | proline-rich pepti |
| 18 | 827.5 | 14.1 | 1344 | 1 A35175 | mucin 1 precursor, |
| 19 | 783.8 | 13.3 | 2232 | 2 T34434 | hypothetical prote |
| 20 | 778 | 13.2 | 2142 | 2 B35098 | MHC class III hist |
| 21 | 777.2 | 13.2 | 5262 | 2 T03454 | ALR protein - huma |
| 22 | 767.9 | 13.0 | 1832 | 2 T31113 | mucin-like glycopr |
| 23 | 766.1 | 13.0 | 3842 | 2 T42730 | Bassoon protein - |
| 24 | 765.9 | 13.0 | 590 | 2 IS4618 | nucleolar phosphop |
| 25 | 765.4 | 13.0 | 761 | 2 C84672 | hypothetical prote |
| 26 | 760.9 | 12.9 | 4006 | 2 T03070 | probable tenascin |
| 27 | 751.9 | 12.8 | 1870 | 2 A43359 | MHC class III hist |
| 28 | 751.2 | 12.8 | 2774 | 2 S43759 | microtubule-associ |
| 29 | 750.9 | 12.8 | 1872 | 2 S35152 | MHC class III hist |

30 749.5 12.7 2225 2 T26063
31 747.7 12.7 4548 1 S00657
32 745.7 12.7 924 2 S27923
33 742.3 12.6 3938 2 T42761
34 732.2 12.4 971 2 T19431
35 723.6 12.3 801 2 T29018
36 723 12.3 1777 2 T34369
37 722.9 12.3 4957 2 T03455
38 722.3 12.3 3381 2 T42389
39 718.2 12.2 1611 2 T38236
40 715 12.1 1630 2 A53577
41 703.7 12.0 3568 1 A40701
42 702.8 11.9 1952 2 T48814
43 700.7 11.9 5170 2 T15348
44 696.5 11.8 3968 2 A44265
45 696.2 11.8 3869 2 A48205

ALIGNMENTS

RESULT 1

A43932
mucin 2 precursor, intestinal - human (fragments)
N:Alternate names: mucin SMUC-41
C:Species: Homo sapiens (man)
C:Date: 10-Mar-1993 #sequence_revision 12-Apr-1996 #text change 09-Jul-2004
C:Accession: A43963; A45106; B45106; B33532; A61257; PQ0328; PQ0329
R:Gum Jr., J.R.; Hicks, J.W.; Toribara, N.W.; Siddiki, B.; Kim, Y.S.
J. Biol. Chem. 269, 2440-2446, 1994
A:Title: Molecular cloning of human intestinal mucin (MUC2) cDNA. Identification of the
A:Reference number: A43963; MUID:94132002; PMID:8300571
A:Accession: A43963
A:Molecule type: mRNA
A:Residues: 1-639 <GU1>
A:Cross-references: UNIPROT:Q02817; UNIPROT:Q14884; GB:L21998
R:Gum Jr., J.R.; Hicks, J.W.; Toribara, N.W.; Rothe, E.M.; Lagace, R.E.; Kim, Y.S.
J. Biol. Chem. 267, 21375-21383, 1992
A:Title: The human MUC2 intestinal mucin has cysteine-rich subdomains located both upstr:
A:Reference number: A45106; MUID:93015075; PMID:1400449
A:Accession: A45106
A:Status: not compared with conceptual translation
A:Molecule type: mRNA
A:Residues: 626-1895 <GU2>
A:Cross-references: GB:M94131; NID:G186395; PIDN:AAA59163.1; PID:G186396
A:Note: sequence extracted from NCBI backbone (NCBIP:116706)
A:Accession: B45106
A:Status: not compared with conceptual translation
A:Molecule type: mRNA
A:Residues: 2037-3020 <GU3>
A:Cross-references: GB:M94132; NID:G186397; PIDN:AAA59164.1; PID:G186398
A:Experimental source: colon
A:Note: sequence extracted from NCBI backbone (NCBIP:116698)
R:Toribara, N.W.; Gum Jr., J.R.; Culhane, P.J.; Lagace, R.E.; Hicks, J.W.; Petersen, G.I
J. Clin. Invest. 88, 1005-1013, 1991
A:Title: MUC-2 human small intestinal mucin gene structure. Repeated arrays and polymor:
A:Reference number: A43932; MUID:91358717; PMID:1885763
A:Accession: A43932
A:Molecule type: DNA
A:Residues: 1343-1350, 'L', 1352-1411, 'S', 1413-1448, 'P', 1450-1503, 'T', 1505-1915 <TOR>
A:Cross-references: GB:M74027; NID:G188863; PIDN:AAA59875.1; PID:G188864
A:Note: sequence inconsistent with the nucleotide translation
A:Note: sequence extracted from NCBI backbone (NCBIN:55749, NCBIP:55750)
R:Gum, J.R.; Byrd, J.C.; Hicks, J.W.; Toribara, N.W.; Lamport, D.T.A.; Kim, Y.S.
J. Biol. Chem. 264, 6480-6487, 1989
A:Title: Molecular cloning of human intestinal mucin cDNAs. Sequence analysis and eviden:
A:Reference number: A33532; MUID:89197956; PMID:2703501
A:Accession: B33532
A:Molecule type: mRNA
A:Residues: 1916-2193 <GU4>
A:Cross-references: GB:M22405; NID:G188873; PIDN:AAA36334.1; PID:G188874
A:Experimental source: intestine
R:Jany, B.H.; Gallup, M.W.; Yan, P.S.; Gum, J.R.; Kim, Y.S.; Basbaum, C.B.

J. Clin. Invest. 87, 77-82, 1991
 A:Title: Human bronchus and intestine express the same mucin gene.
 A:Reference number: A61257; MUID:91086481; PMID:1985113
 A:Accession: A61257
 A>Status: not compared with conceptual translation
 A:Molecule type: mRNA
 A:Residues: 1,1925-1948, 'TTS', 1952-1954 <JAN>
 A:Experimental source: bronchus
 R:Xu, G.; Huan, L.; Khatri, I.; Sajjan, U.S.; McCoil, D.; Jones, C.; Forstner, Biochem. Biophys. Res. Commun. 183, 821-828, 1992
 A:Title: Human intestinal mucin-like protein (MLP) is homologous with rat MLP in the C-b
 A:Reference number: PQ0328; MUID:92198477; PMID:1550588
 A:Accession: PQ0328
 A:Molecule type: protein
 A:Residues: 2328-2468 <XUG>
 A:Cross-references: GB:M85523
 A:Experimental source: small intestine
 A:Accession: PQ0329
 A:Molecule type: protein
 A:Residues: 2328-2342, 'K', 2344-2354 <XUG1>
 C:Genetics:
 A:Gene: GDB:MUC2
 A:Cross-references: GDB:120203; OMIM:158370
 A:Map position: 11p15.5
 C:Superfamily: von Willebrand factor; von Willebrand factor type A repeat homology; von
 C:Keywords: glycoprotein; intestine; tandem repeat
 F:2766-2834/Domain: von Willebrand factor type C repeat homology <VMC>
 Query Match 22.5%; Score 1325; DB 2; Length 3020;
 Best Local Similarity 20.8%; Pred. No. 1.5e-18;
 Matches 417; Conservative 102; Mismatches 500; Indels 990; Gaps 64;

| | | |
|----|------|--|
| QY | 6 | LPYILLLSVVIQVQSODLSSCAGRCG-----EG-----YSR 40 |
| DB | 502 | LQVLAPEVQLFVTLQASQ--GQVQGLCGNPNLEGDDFKTAGLVEATGAGFANWKA 559 |
| QY | 41 | DATCN-----CDY---48 |
| DB | 560 | QSTCHDKLDWDDPCSLAIESANYAEHWCSSLKKTTETPFGRCHSAVDPAEYVKCKYDTC 619 |
| QY | 49 | NCQHYEC---CPDFKRVCTA-----ELSK 71 |
| DB | 620 | NCQNNEDCLCALSSYARACTAGVLMGWREHVCNKDVGSCPNQVFLNLTTCQOTCR 679 |
| QY | 72 | -----GRCFESF-----ERGR-----ECDC-----DAQCKYDK 95 |
| DB | 680 | SUSEADSHCLEGFAPVDCGCCPDHFLDEKGRVCVPLAKSCYHRLGLYLEAGDVVVROER 739 |
| QY | 96 | C-----96 |
| DB | 740 | CVCRDGLHCRQIRLIGQSCTAPKIHMDCSNLATATSKPRALSCQTLAAGYHTECVSG 799 |
| QY | 97 | --CPD-----YESFCAEVHNPSPSSKKAP-----120 |
| DB | 800 | CVCPLMDDDRGCGVKECPVHNNDLYSSGAKIKVDCTCTCKRGWVCTQAVCHGT 859 |
| QY | 121 | -----PPSG-----ASQT 128 |
| DB | 860 | CSIGSGHVIITFDGKYVDPDGHCSYVAVQDYCGQNSLSGFSIITENVPCGTTGVTCSKA 919 |
| QY | 129 | IKSTTKRSPKPNKKTKVIESEB-----ITEKVKNNKRNKKKPTPKPPVVDGAS 182 |
| DB | 920 | IKIFWGRT-ELKLEDKHRVVIQDESGHVAVTREVGQY-----LVVESST 964 |
| QY | 183 | GL-----DNGDFKV-----191 |
| DB | 965 | GIIVWDRKTTVFIKLAPSYKTVCGLCGNFDRSNDDFTTRDMVVSSELDFGNSWKEA 1024 |
| QY | 192 | -TTPDSTTQ-----HNKVSIP-----208 |
| DB | 1025 | PTCPDYSTNPEPCSLNPHRRSWAEKQCSILKSSVFSICHKSVDPKPFVEACVHDSQCDT 1084 |
| QY | 209 | -----KITT 212 |

| | | |
|----|------|---|
| DB | 1085 | GGDCECFCSAVASYAQECTKEGACVFWRTPLCPICFYNNPPHCEWHYBPCGNRSSET 1144 |
| QY | 213 | AKPIN-----PRPSLP- 223 |
| DB | 1145 | CRTINGIHSNISVSYLEGYPCPCPKDRIYEBDLAKCVTADKCGCVEDTHYPGASVPT 1204 |
| QY | 224 | -----PNSDTSKE-----TSLTVNKET 240 |
| DB | 1205 | EETKSCVCTNSQVVRPEEGKILNQDQAGFCWEICGNGTVKEHFNCSITRPSST 1264 |
| QY | 241 | -----TVEKRETTTINKQTSTDGKEKTSK-----266 |
| DB | 1265 | LTFVITLITPTPTSTTTTTTTTTTSSVLSSTFKLCLMSDWINEDHPSGSGDDGRE 1324 |
| QY | 267 | -----ETQSIEKTSKD-----278 |
| DB | 1325 | PDGVCAPEDIECRSVKDPHLSLEQHOGKQVQCVSGFICKNEQDFGNGPFGCYDKI 1384 |
| QY | 279 | -----LAPTSKVLAKPTPKAETTKGPAITTKPEPTTPKBPASTTKPEPTPT 327 |
| DB | 1385 | RVNCWPMKDCITTPSPPTTTTULPPTTPSPPTTTTTPPTTPPTTTTTPPTTPPTTT 1444 |
| QY | 328 | TIKSAP-TTPKBPAPTTTKSAPTTKBPAPTTKBPAPTTKBPAPTTKBPAPTTKSA 386 |
| DB | 1445 | TTTTPLTTTPSPISITTTTTPPTTTPSP-PTTTPSPPTTTPSPPT-TTTTTTPPTTSP 1502 |
| QY | 387 | P-TTP-----KEPAPTTPKBPAPTTKBPAPTTKBPAPTTKBPAPTTKBPAPTT 437 |
| DB | 1503 | PMPTPTTPPASTTTTTPPTTTPSPPTTTTTPPTTTPSPPTTTTTPPTTTPPTTTP 1562 |
| QY | 438 | KEPAPTPKBPAPTTKBPAPTTKBPAPTTKBPAPTTKBPAPTTKBPAPTTKBPAP 497 |
| DB | 1563 | SPPTTTTTPPTTTPSPPTTTTTPPTTTPPTTTPPTTTPPTTTPPTTTPPTTTPSP-P 1621 |
| QY | 498 | TTTKSAPTTKBPSPPTTKBPAPTTKBPAPTTKBPAP-TTPKBPAPTTKBPAPTTTK 556 |
| DB | 1622 | TTT---PITP---PSTTTTTPPTTTPSPPTTTTTPPTTTPSPPTTTTTPPTTTPSP 1674 |
| QY | 557 | KPAPTPKBPAPTTKBPAPTTKBPAPTTKBPAPTTKBPAPTTKBPAPTTKBPAPTT 616 |
| DB | 1675 | PPPTTTPSPPTTTPSPPTTTTTPPTTTPPTTTPPTTTPPTTTPPTTTPPTTTPSP 1725 |
| QY | 617 | TTPKEPAPTTKBPAPTTKBPAPTTKBPAPTTKBPAPTTKBPAPTTKBPAPTTKBP 664 |
| DB | 1726 | TTMTTSPPT-----TPSPPTTTTTPPTTTPSPPTTTTTPPTTTPPTTTPPTTTP 1778 |
| QY | 665 | TTT-----KG-----669 |
| DB | 1779 | TTPCVPLCNWGMDSGKFNHFKPGDTELGVCVPGWAAANISCRATWYDVPVIGQLGQ 1838 |
| QY | 670 | -----TAPTTLKEPAPTTPKKPAP 688 |
| DB | 1839 | TVVCDVSGLICKNEQKPGGVIPMAFCLNYEINVCQCECVTQPTM---TTTTTENPTP 1895 |
| QY | 689 | KELAPT--TKPSTTSIDKP-----APT-TPKGTAPTTKBPAPTTKBPAPTTKBPAP 741 |
| DB | 1896 | TPITTTTTPPTTTPPTTTPSPPTTTTTPPTTTPPTTTPPTTTPPTTTPPTTTPPT 1954 |
| QY | 742 | TTLKEPAPTTKBPAPKELAPTTTKGPTST-TSDPAPTTKBPAPTTKBPAPTTKBPAP 800 |
| DB | 1955 | TVPTPTPTGQTPTTVLITTTTTPPTTTPPTTTPPTTTPPTTTPPTTTPPTTTPPT 2014 |
| QY | 801 | -----APTPPTPTPTSEVSTPTTK-----EPTTIHKSDESTPBLSEPTPKALE 848 |
| DB | 2015 | TMIPISITTTTTPPTTTPPTTTPPTTTPPTTTPPTTTPPTTTPPTTTPPTTTPPT 2074 |
| QY | 849 | NSPKBPPTTKPAPKEMTTAKDKTERDLRTPTTTTAAAPKMTKETATTEKTE 908 |
| DB | 2075 | STTLLSTLPAEMTSTAPPTTATTTSGHLSPPPTTTPPTTTPPTTTPPTTTPPTTTP 2132 |
| QY | 909 | SKTATTTQVSTTQDTPFKITT---LKTTLAPKVTTKTTTITTEIMKNKPEB---961 |

Db 1355 SDEPTSPDETPSDETPSETPBEPT-----PTTTPPTPTPTSGSGSG 1401
 QY 1016 -----TMPELNTSRIAEAMLQTTTRPQNTSKLVEVNPKESEDAGGAEGE 1061
 Db 1402 GSGGGGGGGGTVPTSPFTTTS-----KPTSTPAPTLE-EPFSDVPGAIGG 1449
 QY 1062 TPHMLLRPH---VNPE----- 1075
 Db 1450 EHRAYLRGYPGSGFRPERNITRAEAAVIFAKILGADESYGASAPSYSDLADTHAAWAI 1509
 QY 1076 -----VTPD-----MDYLPVPPNQGI-----IINP 1095
 Db 1510 KPATSQGLFKGYDPDTPFQDONITRAEPATVVLHLFKVKQEIIMSKLATIDISNP 1565

RESULT 3
 T31108
 cyst germination specific acidic repeat protein precursor - Phytophthora infestans
 C:Species: Phytophthora infestans (potato late blight agent)
 C:Date: 22-Oct-1999 #sequence_revision 22-Oct-1999 #text_change 09-Jul-2004
 C:Accession: T31108
 R:Goernhardt, B.
 submitted to the EMBL Data Library, April 1998
 A:Reference number: Z20986
 A:Accession: T31108
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 1-1489 <GOE>
 A:Cross-references: UNIPROT:O96449; EMBL:AF061185; NID:g3851513; PID:g3851514; PIDN:AAC7
 C:Genetics:
 A:Gene: car90

Query Match 17.3%; Score 1020.9; DB 2; Length 1489;
 Best Local Similarity 26.1%; Pred. No. 4.9e-13;
 Matches 406; Conservative 62; Mismatches 425; Indels 661; Gaps 74;

QY 41 DATCNDCVNCQHYMEC-----CPDFKRVCTAELSCCKGRCFBSFERGRCDCDAQ 89
 Db 32 DATYLSLSSCG--IPCSGVGAEPVGTACPKAGDVATS-----DCOPY 71
 QY 90 CKKYD--KCCPYESFCBAVN-----PTSPSSKKA----- 119
 Db 72 LLSYNGAVCVAPDAECALIHDDMMGCEFPKGTGYTSAVEAETIAAYNGESSGWTGHDEV 131
 QY 120 -----PPPSG----- 124
 Db 132 VQVGDEBEI PARVNYDVTDTPIGVNCEVATETATQHAEGGKYVDTPSTGTQTDYG 191
 QY 125 -----ASQIKSTTKRSPKPNKKTKKVISSSE-----ITERKVDNKNRKKKPTTP 172
 Db 192 NTHYGSSTTEGVTGGYGTPTD-----AKVIDETVLDYPTGITIELID----- 235
 QY 173 KPVVDEAGSLDNGDFKVTPTDSTT-----QHNKYSTSKI 210
 Db 236 -----GTPGGYGTITDGGTITGGYTTVDNTHETGAGGYDAGTBEESITPTV 285
 QY 211 -----TTAKPIN-----PRPSLP----- 223
 Db 286 GYSTETEQQVTGGVPSDETEAFTGTYVVPRETTAAPSEDITVAPREVTPVAPTEK 345
 QY 224 -----ENSDT----- 228
 Db 346 PYDVBEITYVBEITYAPTKSTNAPTERMHVAHIEKPCDTEVVMYAPTEETVAPTEET 405
 QY 229 -----SKETSLTVNKEITVE--TKETT--TTNKQSTSDGKEKTSKAKTSQIEKTSKADLAPT 282
 Db 406 TYAPTEITYATEETPYEPTTEITYTPTEITYAPTEETI-----YAPT 450
 QY 283 SKVLAKPTPKAETTKGPAITPKPTPTT-----PKE 315
 Db 451 EKTTVAPT---EETTVAPTEETPYEPTTEITYAPTKETTYAPTEETVASTTEITYAPTE 507
 QY 316 -----PASTTPKEPTPTTIKSAPTTPKEPAPT--TTKSAPTTPKEPAPT--TTKEPAPTTP 367

Db 508 ETTYAPAEETPYEPTTEET--TYAPTBEITYAPTEETTYAPTBEITYAPTEETTYAPAEETP 566
 QY 368 KSPAPTTTTKEPAPTTTTKSAPT-----TPKEPAPTTTTKKPAPTTTTKSPAPTTTTKEP 417
 Db 567 YEPTBEITYAPTEETI--YAPTBEITYAPTEETTYAPTEETTYAPAEETPYEPTTEETVAP 625
 QY 418 TPTTPKEP-----AP-----TPKEPAPTTTTKEP-----APTAKKAPAPTTK 454
 Db 626 TEETTYAPTEETTYASTBEITYAPTEETTYAPAEETPYEPTTEETTYAPTEETTYAPTEET 685
 QY 455 EBAPITPKPAPT--TTKEPSPTTPKEP-----APT--TTKSAPT-----TTKEP 495
 Db 686 TYAPTEETTYATEETTYAPAEETPYEPTTEETTYAPTEETTYAPTEETTYAPTEETTYGP 745
 QY 496 -----APT--TTKSAPT--TPKEPSPTTPKEP-----APTTPKEPAPTTTTK 540
 Db 746 TEETTYAPTEETTYAPTEETTYAPTEETTYEPTTGTETTYAPTEETTYAPTEETTYAPTEET 805
 QY 541 EPAPT--TPKEPAPTTTTKAPAPTAKEP-----APTTPKETAPTTPKLTTP 584
 Db 806 TYAPTEETPYEPTTEETTYAPTEETPYEPTTEETTYTPTEETTYAPTEETTYAPTEETVAP 865
 QY 585 TTPEKLAPT--TPKEP-----APTTPBELAPT-----T 610
 Db 866 TEETTYAPTEETPYEPTTEETTYAPTKETTYAPTEETTYASTBEITYAPTEETTYAPAEET 925
 QY 611 PREPTTPPEEP-----APTTPKAAANTPKAPAPT--TPKEPAPTTTTKEP-----AP 656
 Db 926 PYEPTBEITYAPTEETTYAPTEETTYAPTEETTYAPTEETTYAPAEETPYEPTTEETVAP 985
 QY 657 TTPKETAPTTPKGTAP-----TTLKEPAPTTPKKP-----APKE-----LAPT-- 694
 Db 986 TEETTYAPTEETMYAPIBEITYAPTEETTYAPAEETPYEPTTEETTYAPTEETTYAPTEET 1045
 QY 695 -----TTKEPTSTTSKPA-----PTTPKGTATTPKEPAPTTPKEPAPTTPKGTAP 741
 Db 1046 TYASTEETTYAPTEETTYAPAEETPYEPTTEETTYAPTEETTYAPTEETTYAPTEETTYAP 1105
 QY 742 T--TLKEPAPTTPKEP-----APKE-----LAPT-----TTKGTSTTSKPAPTT 780
 Db 1106 TEETTYAPAEETPYEPTTEETTYAPTEETTYAPTEETTYAPTEETMYAPIBEITYAPTEET 1163
 QY 781 PKETAPT--TPKEPAPTTPKKP-----APTTPPEPTTSEVSTPT--TTKEPTTIHKS 830
 Db 1164 ATTYAPTEETPYEPTTEETTYAPTEETTYAPTEETTYAPTEETTYAPTEETTYAPT--EET 1221
 QY 831 PDSESTBELSAPTPKALENSPKPEGVPTTKTPAATKPEMTTAKDKTTERDLRTTPTT 890
 Db 1222 PYEPTBEITYAPT-----EETTYEPTTEETVAP-----TEETTYAPTEET 1262
 QY 891 AAPKMTKET-----ATT--TEKTESKITATTQVSTTT--TQDPT--PFKITT-- 933
 Db 1263 YAP--TEETMYAPIDETYGVTEETTYAPTEETTYAPTEETTYAPTEETTYEPTTEETTYA 1320
 QY 934 -LKTITLAKVITT-----KKITITTEIMNKPEETAKPKDRATNSKATP 977
 Db 1321 PTEETTYAPTEETTYAPMEETPYEPAESTSTVSTKPCNTEETTEETDEPIDEPTDPSDEP 1380
 QY 978 KPQ--KPTKAPKKPTSTKKKPTMPVRKP-----PCDQNGINGIGVENKVRNAGIYNTTPGR 1429
 Db 1381 TDEPTDEPTDLP-----TDEPST-----SDSVCELATSTSDREEDQCN 1476
 QY 1012 K-----MTSTMPDELNPTSRIAEAMLQTTTRNQTTPN 1042
 Db 1430 NSQSWHSCCRSCVNDPICHAFPHOTS-----SDSVCELATSTSDREEDQCN 1476

RESULT 4
 I38346
 elastic titin - human (fragment)
 C:Species: Homo sapiens (man)
 C:Date: 29-May-1998 #sequence_revision 29-May-1998 #text_change 09-Jul-2004


```
Db 1513 ---EKPA-TPQASGTTASKVPVPAETQEVAVSSRETPTVPAVPPVKNPSSHKKTKTIE 1568
QY 753 KKPAPKELAPTTTKGFTSTTSKPAFTTPKETAPTTPKEPAPTTPKPAFTTPPTT 812
Db 1569 LKEAPATLPPSPKFKIPSSK---APRTSAP
QY 813 SEVSPTTTTKEPTTHKSDESTPE--LSAETPKALENSKEPGVPTTKTAPAKPMT 870
Db 1613 SLAQT---APPSLQKAPTTPKENLAA---PAVLVSSKSPAAP-ARASASLSP--- 1660
QY 871 TTAKDKTTERDURTP-----ETTTA-----APKMTKATTTKTTESKITA 913
Db 1661 ATAAPQATPEKATTPSCKAAATPTIETSTAPLEGAPKETSTSV-----SKYLM 1713
QY 914 TTTQVTSITTQDTPFKITLL---KTTTLPKVTITTKTITTEIMNKPEETAKPKDRA 969
Db 1714 SPPKKAASSKRASTLPATLPSLKEASVLSSTATSSGK-----DSHISPVSDA 1762
QY 970 TNSKATTPKPKQPTKAPKX-----ETTTA-----APKMTKATTTKTTESKITA 988
Db 1763 CSTGTTT--POASEKLPSKKGTAFTTEMLAAPAPBSALAITAPIQKSPGANSASSPKC 1820
QY 989 --PTSTKPKTMR--VRKPKTTPTRKWTSTMPNLNPTSRIAEAMLOTTPRPNQ----- 1039
Db 1821 PPSKSKOTKGLPSAVALAPQTVPEK-----DTSKAIETLLVSPAKGSDCLHSP 1870
QY 1040 -----TPNSKLVEVNPKSEADAG 1056
Db 1871 KGPVGSQVATPLAAFTSDKVPPEAVSASVAPKAPAPASLITLAPSPVAPLPKQPILLESAP 1930
QY 1057 GAEGTTPHMLRPHVMEVTPDMYLRV-----PNQGIHIN----- 1094
Db 1931 GSVLESPSKL-----FVPAEDELPLPPEAVSGGEFPQPLVNMMPAKPAGTPAP 1982
QY 1095 -----PML 1097
Db 1983 APSAKQPV 1991

RESULT 7
T45025
mucin MUC5B, tracheobronchial [imported] - human (fragment)
C:Species: Homo sapiens (man)
C>Date: 21-Jan-2000 #sequence_revision 21-Jan-2000 #text_change 21-Jul-2000
C:Accession: T45025
R:Desseyn, J.L.; Guyonnet-Duperrat, V.; Porchet, N.; Aubert, J.P.; Laine, A.
J. Biol. Chem. 272, 3168-3178, 1997
A:Title: Human mucin gene MUC5B, the 10.7 kb large central exon encodes various alternat
A:Reference number: 222899; MUID:97166151; PMID:9013550
A:Accession: T45025
A>Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-3570 <DES>
A:Cross-references: EMBL:Z72496; NID:g1834502; PIDN:CAA96577.1; PID:g1834503
A:Experimental source: placenta
C:Genetics:
A:Gene: MUC5B

Query Match 16.9%; Score 992.1; DB 2; Length 3570;
Best Local Similarity 13.1%; Pred. No. 5.9e-12;
Matches 444; Conservative 126; Mismatches 459; Indels 2359; Gaps 103;
QY 23 SSQDLSSCAGRCG---EGYS-----RDATC----- 44
Db 172 SAFGTTTCQPCQWTEWDEDEYKSEQLGGDVESYDKIRAGGHLCCQPKDIEQAESFP 231
QY 45 -----NCD-----YNCQHYMECCPD----- 59
Db 232 NWTLAQGVQKCHVDHFGVLGVCNWEQGVFKMVCNRYRVLVCCSDDHCRGATTPPTTE 291
QY 60 ----- 59
```

```
Db 292 LETATTTTQALFSTPQPTSSPOLTRAPPASTTAVPTLSEGLTSPRYTSTLGTATTGGPR 351
QY 60 ----- 59
Db 352 QSAGSTEPTVPGVATSTLPTRSALPGTGSGLGWPRSPQPTLAPTMTATSRARPTGTAST 411
QY 60 -----FKRYCTAELS-----CKGRC----- 74
Db 412 ASKEPLTSLAPLTSELSTQAETSTPRTEITMSPLTNTTTSQGTTRCQPKCEWTFD 471
QY 75 -----FESPER-----GREC-----DCDAQ----- 89
Db 472 VDFPTSGVAGDMETPENIRAAAGKCMWAPKSIIECAENYPEVSDQGVLTCSLETGL 531
QY 90 -CKYDK-----CCPDYESFC-----AEVHNPTSP----- 114
Db 532 TCKNEDOTGFNMCNPNVNVLCDDY-SHCPSLTATSTATSPSTPGTWTWILTKPTITA 590
QY 115 -----SSKAPP----- 121
Db 591 TTTASTGSTATASSTOATAGTAPHVSTATTPTVTSKATPFSSPGTATPALRSTATTP 650
QY 122 -----PSGASOTI----- 129
Db 651 TATSFTALPSSSLGTTWRLSQTTTPMATMSTATSSSTPETVHTSNVLTATTATTGATGS 710
QY 130 ----- 129
Db 711 VATPSTPGTAHTKVLTTTGTATPSSSPGRATLFPWISTTTTPTTRGSTVTPSSI 770
QY 130 -----KSTTKRSPKPNKKTKKVIESBEITEKVDKNKN 164
Db 771 PGTHPTVLTVTITTTVATGSMATPSSSTQTSGLTPSLTTTATTATTAGST----- 821
QY 165 RTKKKP-----TPKPPV----- 177
Db 822 ---TNFSSPTGTTPIPPVLTATTATPAATSTVTPSSALGTTTTPPVNNTATTGRLS 878
QY 178 -----DEAGSLDNGDFKVTTP-----TSTQHNKVSTSPKI----- 210
Db 879 PPSPTVCTAWTATSGLI-LGTHITEPSTGTSHTPAATTGTQH-----STPALSSPHS 933
QY 211 --TTAKPINPRPSLPNSDTSKETSITVNKE-----TTV----- 242
Db 934 SRTESPPSPGTTTPGHTTATSTTATATPSKTRTSTLLPSQTSAPITVVTMGEPCQ 993
QY 243 ----- 242
Db 994 AMSEWLDYSYPMGPGSGDPTYSNIRAAAGAVCEQPLGLECRAQAQPGVRLQGVVE 1053
QY 243 -----ETKETTNTNKQSTDG-----KE 260
Db 1054 CSLDFGLVCNREOVCKFKMVCNVEIRVPCNMGHCPSPTATSSSTATSSSTGTTWILTE 1113
QY 261 KTSKAKETSIEKTSKADLAPSKVLAKPTPKAETTKG----- 299
Db 1114 LTTTATTTSTGSTA---TPSTLRTAPPKVLATTATTPTVTSKATPFSSPGTAT 1169
QY 300 PAL-----TTP-----KEPPT-----TPK----- 314
Db 1170 PALRSTATTPTATSVTPPISSSLGTTWRLSQTTTPTATMSTATSSSTPETATHTSVLTA 1229
QY 315 -----EPAST-----TPKEPTPTTI-----KSAPTTPKEP-----APT 343
Db 1230 TATTGATGATVATPSPSTPGTAHTTKVPTTTTGTATPSSSPGTALTPPWISTTTTPT 1289
QY 344 -----TKSAPTTPK----- 352
Db 1290 RGSVTVPSSIPGTHTATVLTITTTTATGSMATPSSSTQTSGLTPSLTTTATTATTGTS 1349
QY 353 -----EPAP-----TTTKEPAPT-----TPKBPAPTTTKEPAPTTKS-----APT 388
Db 1350 TTNPSSTPGTRPIPPVLTITTTTATPAATSTVTPSSALGTTTTPPVNTATTGRLSPS 1409
```

QY 389 TP-----KEPAPTTKKAPATT-----PKEPAPTTKKE 416
Db 1410 SEHTVRAMTSATSGTLGTHITEPSTGTSHTPAATTGTHQSTPALSSPHSPSTTESP 1469
QY 417 PTP--TTPKEPAPTTKEPAPTTKE----- 439
Db 1470 PSFGITTPGHTTATSRTTATATPSKTRTSLTPSSFTSAPITTVTMGCEPOCAWSEWLD 1529
QY 440 -----PAPTA-----PKKP----- 448
Db 1530 YSEMPGSGGDFDTYSNIRAAAGAVCEQPLGLECRAQAQGVPLRELQVQVECSLDRL 1589
QY 449 -----APTPKKEPAPTTKE----- 463
Db 1590 VCRNEQVGFKMCFNVEIRVFCNNGHCPSTPATSSTATPSSTPGTTWILTEQTAATT 1649
QY 464 -----PAPTTTKEPSTTPK----- 478
Db 1650 TATTGSTALPSSTPGAPPPKVLTSQAATPTATSSKATSSSPRTATLPLVLTSTATKST 1709
QY 479 ----- 478
Db 1710 AFSFTPIPSSTLGTGTSQNRPPHPMATWSTIHPSTPETHTSTVLTKATTRATSSM 1769
QY 479 -----EPAPTTTKSAPT----- 490
Db 1770 STPSSTPGTTWILTELTTAAATTAALPHGTPSPSTPGTTWILTEPSTTATVTPGSTAT 1829
QY 491 ----- 490
Db 1830 SSTRATAGTLKVLTSATPTVISSRATPSSPGTATALPALRSTATTPTATSVTAIPSS 1889
QY 491 ----- 490
Db 1890 SLGTAWRLSQTTTPATMSTATPSSTPETVHTSVLTATTTATTRGVSATPSSTPGTAH 1949
QY 491 TTKEPAPTTT-----KSAPTTKPEP-----SPTT----- 514
Db 1950 TUKVPTTTTGTFTATSSSGHALTPPWISITTTPTTRGVSVPSSIPGTHHTAVLTT 2009
QY 515 -----TKEPAPTTKPEP-----TTPK 532
Db 2010 TTTTATGSMATPSSSTQSGTPPSLTATTATITATGSTTNPSSTPGTTPIPVLTATTAT 2069
QY 533 KPAPT-----TPKEPAPTTTKEPAPTTT----- 555
Db 2070 TPAATSSVTPSSALGTHTPPVNTTATTHGRSLPPSPHVPVPAWTSATSGILGTHI 2129
QY 556 -----KKAPT-----APKEPAPTT-----KETAPTT 578
Db 2130 TEPSTGTSHTPAATTGTTQFSTPALSSPHSPSTTESPPSPGTTTPGHTRGTSRTATAT 2189
QY 579 PKK-----LTPTTPEKL----- 590
Db 2190 PSKTRTSLPSSFTSAPITTVTTCCEPOCAWSEWLDYSYMPGPGSGGDFDTYSNIRAA 2249
QY 591 ----- 590
Db 2250 GGAVCEQPLGLECRAQAQGVPLRELQVQVECSLDPLGLVCRNREQVGFKMCFNVEIRVF 2309
QY 591 -----APTPKKEPAPTTPEELAPT-----PEPTP- 616
Db 2310 CNYGHCPSPTATSTATPSSTPGTTWILTKLTTTATTTTESTGSTATPSSTQGPAGTGH 2369
QY 617 --TTPKEPAPTTPKA----- 629
Db 2370 VSTTATTPVTSSKATPSSPGTATALPALRSTATTPTATSTFATPSSSLGTTWRLSQT 2429
QY 630 -----AANTPKKEPAPT-----TPKEPAPT--TPKEPAPTTPK 660
Db 2430 TTPMATMSTATSSTPETVHTSTVLTTTATTGATGVSATPSSTPGTAHTTKVPTTTTIG 2489

QY 561 ET-APTTPKGA-----PTT-----LK 676
Db 2490 FTVTPSSPGTARTPPWISITTTPTTSGSVTPSSIPGTHHTPVLVLTTPQPVATCSMA 2549
QY 677 EPAPTTTTPKAPKELAPT-----TTKEPTSTTSKAP-----TTPKGTAPT-TP 720
Db 2550 TSPSSSTQSGTPPSLIITATTATGSTTNPSSTPGTTPPELTATTTATPAATSSVTP 2609
QY 721 KEAPATTTPKBPAPT----- 736
Db 2610 SSALGTHHTPPVNTTATTHGRSLPSSPHVTARTAMTSATSGTLGTHITEPSTGTSHTP 2669
QY 737 KGTAAPT-----LKBPAP-----TTPKAPKELAPTTTKGTSTTSKAPTPKETAPT 787
Db 2670 AATTGTTTTPALSSPHSPSTTESPPSP-----GTHTPGHTTATSTTATPAKSTRTS 2725
QY 788 T--PKEP--APT----- 796
Db 2726 TLLPSQTSAPITTVTTCCEPOCAWSEWLDYSYMPGPGSGGDFDTYSNIRAAAGAVCEQ 2785
QY 797 -----PKKP-----A 801
Db 2786 PLGLECRATAQGVPLGELQVVECSLDPLGLVCRNREQVGFKMCFNVEIRVFCNNGHC 2845
QY 802 PTTPET-----PPP--TTSVST 817
Db 2846 PSTATSSATPSSSTPGTTWILTELTTATTATGSTATSPSTPGTAPPKVLTSAPT 2905
QY 818 PTTTKEPTTHKSP----- 831
Db 2906 PTATSSKATSSSPRTATLPLVLTSTATKSTATSVTPIPSTLGTGTLPEQITTPVATM 2965
QY 832 -----DESTPE-----LSAEPKALENGPKPGVP-----TTK 860
Db 2966 STIHPSTPETHTSTVLTKATTRATSSSTPSSSTPGTTWILTELTTATTAATGTOPTA 3025
QY 861 TPAATK-----PEMTTAA----- 873
Db 3026 TSPSTPGTTWILTELTTATTATGSTATSSSTPGTTWILTEPSTTATVAPPGSTAT 3085
QY 874 -----KDKTTE-----RDLRTTPEIT----- 889
Db 3086 SSTQATAGTHVSTTATTPVTSKATPSSSPGTATAPALRSTATTPTATSTFATIPSSS 3145
QY 890 -----TAAPKMTKETA-----TTTEKTT----- 907
Db 3146 LGTTWRLSQTTTPTATMSTATPSSTPETVHTSVLTATTATTGATGVSATPSSTPGTAH 3205
QY 908 ESKITATTG-----VTSITQDTP-----F 929
Db 3206 TTKVPTTTTGTFTATPSSSPGTALTPPWISITTTPTTPTTSGSVTPSSIPGTHHTA 3265
QY 930 KITTLKTTILA-----PKVTTTKITITTEIMNKPEETAKPKDRATNS 972
Db 3266 RVLTTTITTVATGSMATPSSSTQSGTPPSLTATTATITATGSTTNPSST----- 3315
QY 973 KATTPKP-----QKTPKAPKPTSTKPK-----TMPVRK----- 1003
Db 3316 PGTTPIPVLTSMATTPAATSSKATSSSPRTATLPLVLTSTATKSTATSTFATIPSSTLW 3375
QY 1004 -----PKTTPTPKX-----TSTMPELNPTSRIAEAMLQTT-----TRNCTPNSKLVE 1047
Db 3376 TWTWVPAQTTPMTSTMSTIHTSSTPETHTSTV-----LTATMTATINATIPSSLTGT 3431
QY 1048 VNPKE-----DAGGAFETPHMLLRP-----H 1070
Db 3432 TRIULTELTATTATAAGSTATSSSTPGTTWILTEPSTIATVMVPTGSTATSSLTGTAH 3491
QY 1071 VFMVEVTPDXYLP-----XVBNQGI 1092
Db 3492 T--PKVVTANATMPTATATAPSSSTV 3516

RESULT 8
S48478
Glucan 1,4-alpha-glucosidase (EC 3.2.1.3) - yeast [Saccharomyces cerevisiae]
N:Alternate names: extracellular glucosylase; mucin-like protein MUC1; protein YIR019c
C:Species: Saccharomyces cerevisiae
C:Date: 10-Sep-1999 #sequence revision 10-Sep-1999 #text_change 09-Jul-2004
C:Accession: S48478; B26877; S27281; JG6123
R:Rowley, K.
submitted to the EMBL Data Library, October 1994
A:Reference number: S48478
A:Accession: S48478
A:Molecule type: DNA
A:Residues: 1-1367 <ROW>
A:Cross-references: UNIPROT:P08640; GB:247047; EMBL:Z38061; NID:G603997; PID:G763364; GS
R:Yamashita, I.; Nakamura, M.; Fukui, S.
J. Bacteriol. 169, 2142-2149, 1987
A:Title: Gene fusion is a possible mechanism underlying the evolution of STA1.
A:Reference number: A91831; MUID:87194600; PMID:3108330
A:Accession: A26877
A:Molecule type: DNA
A:Residues: 1-242 <YAM>
A:Cross-references: EMBL:M16164; NID:G172522; PIDN:AAA35014.1; PID:G172525
A:Accession: B26877
A:Molecule type: DNA
A:Residues: 762-1331 <YA2>
A:Cross-references: EMBL:M16165; NID:G172523; PIDN:AAA35015.1; PID:G172526
F:Rado, J.M.; Ianez, E.; Zalacain, M.; Claros, M.G.; Jimenez, A.
FEBS Lett. 239, 179-184, 1988
A:Title: Similar short elements in the 5' regions of the STA2 and SGA genes from Sacchar
A:Reference number: S27281; MUID:89031230; PMID:3141213
A:Accession: S27281
A:Molecule type: DNA
A:Residues: 1-31 <PAR>
A:Cross-references: EMBL:X13857; NID:G4551; PIDN:CAA32069.1; PID:G4552
R:Lambrechts, M.G.; Bauer, F.F.; Marmur, J.; Pretorius, I.S.
Proc. Natl. Acad. Sci. U.S.A. 93, 8419-8424, 1996
A:Title: Muc1, a mucin-like protein that is regulated by Mss10, is critical for pseudohy
A:Reference number: JG6123; MUID:96323237; PMID:8710886
A:Accession: JG6123
A:Status: nucleic acid sequence not shown
A:Molecule type: DNA
A:Residues: 1-1367 <LAM>
A:Cross-references: GB:U30626; NID:G1304386; PIDN:AAC49609.1; PID:G1304387
C:Genetics:
A:Gene: SGD:MUC1; STA2; MAL5; DEX2; SGD:S0001458
A:Cross-references: MIPS:YIR019c; SGD:S0001458
A:Map position: 9R
C:Superfamily: yeast glucan 1,4-alpha-glucosidase homolog; glucan 1,4-alpha-glucosidase
C:Keywords: glycosidase; hydrolase; polysaccharide degradation; transmembrane protein
F:5-21/Domain: transmembrane #status predicted <TM1>
F:1350-1366/Domain: transmembrane #status predicted <TM2>

Query Match 15.7%; Score 923.9; DB 1; Length 1367;
Best Local Similarity 24.6%; Pred. NO. 3.4e-11;
Matches 341; Conservative 113; Mismatches 426; Indels 506; Gaps 69;

| | | | |
|----|-----|--|-----|
| QY | 5 | TLPIVLLLLSVFVYIQVSSODLSCAGRCGEGYSDATCNDYNCQHYMECCPDFKRVK | 64 |
| DB | 129 | TFEYVY-----ATQDVNS-----CQVWM--PWFQIQ- | 151 |
| QY | 65 | TAEISCKGRCF-----SFERGEC-----DC | 86 |
| DB | 152 | -----FEYLOCSAAQYASSQWGTGTFDLSTGCNVDNQHSGTDPFGFYMNIDC | 201 |
| QY | 87 | DAQC----- | 100 |
| DB | 202 | DNCCGCTKSSITTSSTSSSTTSSTSSSTTSSTSSSTTSSTSSSTTSSTSSSTTSSTAP-- | 259 |
| QY | 101 | BSFCAEHNPTSPSSKAPPSPGASQTIKSTKSPKPPN-----KKTKKVIKSEE | 153 |
| DB | 260 | -----ATPTTSCTEKPTP-----PTTSCTEKPTPPTPHDTPCTKKT----- | 300 |
| QY | 154 | ITEKYDKNKNRFTKKKPTPKPVVDAGSLDNGDFKVTPTDSTTQHNKYSTSPKLTIA | 213 |

| | | | |
|----|------|--|------|
| DB | 301 | -----TTSKTCIKKITTVP-----TPSSSTTE-----SSAPVPT- | 331 |
| QY | 214 | KPINPRPSLPNDSDTSKESLVINKETT-----VETKETTINKQTS--TWGKKTISA | 265 |
| DB | 332 | -----PSSSTTESSAPVTSSTTESSAPVTPSSSTTESSAPVTSSTTESSA | 381 |
| QY | 266 | KETQSIKTSAXDLAPTCKLAKPTKARTTKGPA-----LTPKEPTTPKPEASTTP | 321 |
| DB | 382 | PVTSSTTESS-----APV-----PTPSSSTTESSAPVTSSTTESSAPVT-----SSITE | 428 |
| QY | 322 | KEPTPTTI-----KSAP-----TTPKEPAPTTTKSAPTTTKEPAPTTTKEPAPTTTKEPAP | 372 |
| DB | 429 | SSAPVTSSTTESSAPVTSSTTESSAPVTPSSSTTESSAPVTS-----TTSSSAP | 484 |
| QY | 373 | TTPKEPAPTTKSAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAP | 422 |
| DB | 485 | VTPSSSTTESSAPVT-----SSTTESSAPVTPSSSTTESSAPVT-----PTPSSSTTE | 536 |
| QY | 423 | KEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAP | 473 |
| DB | 537 | SSAPVT-----SSTTESSAPV-----PTPSSSTTESSAPVTPSSSTTESSAPVTPS--S | 586 |
| QY | 474 | PTTPKEPAPTTKSAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAP | 532 |
| DB | 587 | STTESSAPVTPSSSTTESSAPVTPSSSTTESSAPVTPSSSTTESSAPVTPS-----TTES | 642 |
| QY | 533 | KPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAP | 592 |
| DB | 643 | SSITTESSAPVT-----PSSSTTESSAPVTPSSSTTESSAPVT-----SSTTESSAP | 694 |
| QY | 593 | TTPKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAP | 643 |
| DB | 695 | VT-----SSTTESSAPV-----PTPSSSTTESSAPVTPSSSTTESSAPVTPSSSTTE | 746 |
| QY | 644 | KEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAP | 702 |
| DB | 747 | SSAPVT-----SSTTESSAPVTPSSSTTESSAPVTPSSSTTESSAPVT-----PSS | 800 |
| QY | 703 | TSDK-----PAPTTKGPATTPKEPAPTT-----TPKEPAPTTKGPAT-----TT | 743 |
| DB | 801 | TESSVAPVTPSSSNTSSAPSTPSSSTTESSAPVTPSSSTTESSAPVTPSSSTTESSAPVSS | 859 |
| QY | 744 | LKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAP | 798 |
| DB | 860 | EGSVAPV-----PTPSSSNTSSAPSSIPFSSITTSFSTGTT---VTPSSSKYPSQTH | 912 |
| QY | 799 | KPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAP | 856 |
| DB | 913 | SVSSTETTVIPKTTTSVTPSTTTTTCSTGNSAGETTSKGPVTTT-----V | 967 |
| QY | 857 | PTTKPAATKPEMT-----TAKDKTERDLRTPET-TTAAKPMKETAATTEKTE | 908 |
| DB | 968 | PTTTSVTSTSTTTTTCSTGNSAGETTSKGPVTTTTCSTGNSAGETTSKGPVTTTTCSTG | 1027 |
| QY | 909 | SKITATTTQVIST--TTQDTTTPFK-----ITLKTTLA--PKVTT-----TK | 947 |
| DB | 1028 | SPTPVTVVSVTVVTVVTVVTVVTVVTVVTVVTVVTVVTVVTVVTVVTVVTVVTVVTVV | 1087 |
| QY | 948 | KITTEI---MNKPEET---AKPK----- | 966 |
| DB | 1088 | TTTTCSTGNSAGETTSKGPVTTTTCSTGNSAGETTSKGPVTTTTCSTGNSAGETTSKGP | 1147 |
| QY | 967 | -DRATNS--KAT-----PKQPTKAPKPF----- | 989 |
| DB | 1148 | SSTGNSAGETTSKGPVTTTTCSTGNSAGETTSKGPVTTTTCSTGNSAGETTSKGPVTTT | 1207 |
| QY | 990 | TSTKPKPTM-----PRVKPKPTTTPRKMKTSTMPELN-----P | 1022 |
| DB | 1208 | TSVCSAKTIVSSASAGENTAPSATTPVTTAIPTTVTTTSSVGTNSAGETTTGYTTKSI | 1267 |
| QY | 1023 | TSRIAEAM-----LQTTTRPNOTPNSKLVNPKSEDAGAGETPHMLLRHV | 1071 |

Db 1268 TTYITTLIPSGNAGNYETVATNP-----ISIKTTSQLATASASS-----VAPVW 1315

Qy 1072 FMPEVT 1077

Db 1316 TSPSLT 1321

RESULT 9

T34513

hypothetical protein ZK783.1 - Caenorhabditis elegans

C:Species: Caenorhabditis elegans

C:Date: 29-Oct-1999 #sequence_revision 29-Oct-1999 #text_change 09-Jul-2004

C:Accession: T34513

R:Favallo, A.; Vaudin, M.

submitted to the EMBL Data Library, August 1994

A:Description: The sequence of C. elegans cosmid ZK783.

A:Reference number: Z21536

A:Accession: T34513

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 13507 <FAV>

A:Cross-references: UNIPROT:Q23587; EMBL:U13646; PIDN:AAC24418.1; GSPDB:GN00021; CESP:ZK783

A:Experimental source: strain Bristol N2; clone ZK783

C:Genetics:

A:Gene: CESP:ZK783.1

A:Map position: 3

A:Introns: 14/1; 48/2; 84/1; 196/3; 303/1; 381/1; 586/1; 605/1; 1175/3; 1207/1; 1409/2; 3504/1

Query Match 15.5%; Score 910.5; DB 2; Length 3507;

Best Local Similarity 13.3%; Pred. No. 2.3e-10;

Matches 405; Conservative 183; Mismatches 448; Indels 2005; Gaps 94;

Qy 7 PIVLLLLSVFVIQVSS-----QDL-----SSC 30

Db 32 PTFVVPDTSVICOHSSDDTDLIHNMSSLDGKQDCFCVNPAMHDEVFPYCEHKQSTC 91

Qy 31 AGR-----CGEGYSRDA-----TCNC-----46

Db 92 SGRGACLYDCAKPCQCYDSGFSACBLODKNELEHPCHEMAQCONTLGSVECRCLPGY 151

Qy 47 ---DYNCQHWEC-----CPDFKRV-----63

Db 152 EGNHECTDIDECSDKLTSCRPHSKCINLPGYNYCCTGFTPKNGQSLDKCADINE 211

Qy 64 -----CTAELSCK-----GRCFESFERGRE-----83

Db 212 CETGAHNCDADEICENSIGSFKNKCPGYELIDGKCEDVNECGSEKHLKCDVRADCVN 271

Qy 84 -----CDC-----86

Db 272 TIGGYECEBEGPEGDKNCQKQKSHSVKLPKSKCRKNKSAICDRHASCHIVLDICDCKT 331

Qy 87 -----DAOCK-----92

Db 332 GYTGDGITCHDINECDADTPCSDDGRCNLNLDGGYVCCNKQDDATCIRKQGAFCSGGG 391

Qy 93 -----YDKC-----CPDYESFCAEV-----HNPTS 112

Db 392 DNAICSNATCACIDGFRGDPHKVCVDINECVENDSVCGVGDRCVNLFGGFKCCQHGSTE 451

Qy 113 PPSKKA-----PPPSASQTIKSTTKRSPK-----138

Db 452 AECDQAPSSDSSSTISSHGADFTTGEQIIEGSGSIQTSSGSLTV--TRGLIPKQVELT 509

Qy 139 -----PPN-----141

Db 510 TSGRLACTSYCFPNSECVGVCYGVGNALVGCEDIDECITEICNIEANECVNLIG 569

Qy 142 -----KKTKKVIESEE-----153

Db 570 GFVCCNPNTNATHDDCIDFLTKVKIYAYMIIFLLKGLLEITKEKGLHVGNEEDTVVATR 629

Qy 154 -----ITEKVDKNKURTKK-----168

Db 630 SNHSTSDQLITQVVQSRNFSTGOIILTRCKVSSGEAVTQTTDADBFGLISAADLAGSG 689

Qy 169 ----KPTPKPPVDEAGSL-----DNGD-----188

Db 690 SGITLPTTLPEPKIE--GSGKASGVWTEDEGEDEDLMEEGSGSWTTINGITGSPR 747

Qy 189 ----FKVTT-----PDTSTT-----199

Db 748 SEGIRVRIITLGDGEPEFATKPGISAPDKTCGSKTESDGEKLTVEKDGEAQSOGS 807

Qy 200 -----QHNKVSTSPKI-----TTA 213

Db 808 SATSSGKKSEATSGSSSSSAKSGTSEASGSSGASSSSGSGVSGSGSSVSTESGSGFTS 867

Qy 200 -----QHNKVSTSPKI-----TTA 213

Db 868 SSGSVGEATGTVDGSESGKPSKSTHEKLPFTKNGEKSPISGSDTTGKSESEETSR 927

Qy 214 KPINPRPSLPENS-----DTSKETSLTVNK-----238

Db 928 KPFGESDLSLTGSGSGGEWFTGSKGHFESGSKVSTGKGTQSGAEGSGGPKVPKPG 987

Qy 239 ---ETTVEKETT---NKQSTDCGKETTSAKETCSI 271

Db 988 APEITTDGEESSTGDKGCKPADKSDKNKVPKTDGDKNPDIITDGEDSTSETSGGEQG 1047

Qy 272 EKTSAKLAP-----TSKVLAKPT-----290

Db 1048 PKGSKGQPPGDKGSEVKKPTSEVDGPGNLGKSNVPLKPTDLPPEGSGILTSSGG 1107

Qy 291 ----AETTKGPAITPKPEPTT---TPKE-----PA 317

Db 1108 KNSTPEHGTGLERLPKPTEDKSETPOLGLEISAKKPEPEDGTSKEVLEILWESTTPG 1167

Qy 318 STT-----320

Db 1168 STTLDSDSVGLIEISGDLTKATKPHVIEGSGTGDEEITATTRDVSKSTKKPRVEVDGG 1227

Qy 321 ----PKPEPTTIKSA-----PTT-----PKPE-----339

Db 1228 DNGETSGVDGKPTTAPTPSSASBSTSRIPTTSEASPEGSGGAGVPEPDGSGESSTS 1287

Qy 340 ----APTTKSAPTTPKEPAPT---TKEPAPTTTPKEPAPT---374

Db 1288 APDGVSPSTSAPEVPTTSASSTPDVAEESGIFSTSKPTAEPLTT---APTEVTSP 1343

Qy 375 ----TKEPAPTTKSAPTTPKEPAPT---KKPAPTTTPKEPAPT---413

Db 1344 EGSGETEELPTPEGSGESTTSSAPT-EPATVLPQNRNEKPEPTKDTALPTTTTGA 1401

Qy 414 ----PKE-----416

Db 1402 PQANDSSVENTKCTSSDEGLDALCERRTGVCRCPEGAPPKKSCVDVDECATGDHNC 1461

Qy 417 ----PT-----418

Db 1462 HESARCONVGVGYACFCPTGFRKADGSCQDIDECTEHNSTCCGANAKVKNKPGYSCEC 1521

Qy 419 ----PTTPKEPAPTTK-----430

Db 1522 ENGFLGDGVQCVPTT-KKPCDSTQSSKSHCSNMSCSVDTVDGSVECKEKGKYSK 1580

Qy 431 ----EPAP-----434

Db 1581 VCEDINECVAEKAPCSLNCANVMNGTFCSCCKQGYRGDGMCTDINECDERHPCHPAE 1640

Qy 435 ----TPKE-----439

Db 1641 CTNLEGSFKCECHSGFEGDGIKKCTNPLERSCEDVEKFCGRVDHVSCLSVRIYNGSLSV 1700

Qy 440 ----PA-----441

Db 1442 KKEKSSPSPTTKTGDEVEKSP-----PKSP-----TKKESPEKPEDVKSPVKKEKS 1489
QY 372 PTTT-----KEPAPTTTKGAPTT----- 390
Db 1490 PDAATNIVESSSETTIEKTTTMTTTHESBSRSVVKKEKTEPKVDEKPKSPTKKDKSP 1549
QY 391 -----KEPAP--TTPKKAPATTPKGPATTPKPEPTTPKPE--PAPTTPKEPAP-- 434
Db 1550 EKSITBEIKSPVKKEKSPKVEKAPASPTKKEKSPKASPTKKEGENEVKSPKKEKSPKE 1609
QY 435 -----TTPKGPAP--TAPKKAPATTPKGPATTPKPEPAPTTTPKPEPAPTTTPKPEPAPTTTK 486
Db 1610 KSVVEELKSPKSPKADDDKPSPTKKE-----KSPKESATKEDVKSPKKEKSPKVE 1663
QY 487 SAPTTTKEPAPTTTKSAPT-----TPKPEPPTTKE--PAPTTPKEPAP----- 528
Db 1664 EKPT-----SPTKKESSPTTKTDDEVKSPKKEKSPQIVEEKPASPPTKKEKSPKESKSPKSVVEE 1718
QY 529 -TTPKKPAPTKE--PAPTTPKEPAP-----TTTTKK-----PAPTAPKGPAPTTPKET 574
Db 1719 VKSPEKSPKAEKPKSTKKEKSPKESKAAEVKSPKKEKSPKESKAAEVKSPKKEKSPKES 1778
QY 575 APT-----TPKKL-----TPTTPEKLAPTTPKPAPTTPPELAPTTPPEP 614
Db 1779 SPVQMAADDEVKSPKKEKSPKVEKAPSTKKEK-----TPEKSA--ABELKSPKKEK 1831
QY 615 TPTTP-----EBAPTTPKAAANTPKPEPAPTTPKPEPAPTTPKPEPAP-- 656
Db 1832 SPSSPTTKTGBSKEKSPKPEKPKSPTKPKSPFGSPKK-----KKSKEPEAKPPAPK 1886
QY 657 ----- 656
Db 1887 LTRDLKLTQVNTKDLAHPVVVVEHATECKWFLDQKEITTAQGVTVSKDDQPEFRCSIDTT 1946
QY 657 -----TTPKET----- 662
Db 1947 MFGSGTVSVVAGNAGSVETKELVLETPEKTEKPEFTDKLRDMEVTKGDTQVQMDVIAL 2006
QY 663 -----A 663
Db 2007 HSPLXWYQNGNLLBEGKNGVTKIENKSSLIIPNAQDSKLIIVASNEVGSSESAQL 2066
QY 664 PTPKGTAPTTLKEPAPTTPKKAPKAPKELAPTTPKPTSTTSKDPAPTT----- 711
Db 2067 TVNPPSTTPIVVDGPKSVTIKE-----TETASFKATISGFAPPTVKWTINEKIVEES 2118
QY 712 -----PKGTAPTTPK-----EPAPTTPK----- 729
Db 2119 RTITTIKTEVDVYTLKISNAKIBQGTGVKVTAGNSAGQSKQADLKVEPNVAPKPKSQLT 2178
QY 730 -----EPAPTTP-----KGAPTTLKEPAP-- 749
Db 2179 DKVADEGEPLRWNLDELGSPGTEVSVLLNGQPLKSDTVQVWDGDTGVYHTIAEAKPE 2238
QY 750 -----TTPK-----KPAP 757
Db 2239 MSGTLTAKANAAAGCE--SAKVTVNGNKKPEFVQAPQNHETTLSESVKFAIVTGKPMF 2298
QY 758 K----- 758
Db 2299 NVTWYLNKKLIQSEEVKVKYVHETGKT--SIRIQPLMEHNGTIRVEAENVSGKVAQAL 2358
QY 759 -----ELAPTTT-----KGPTSTTSK 775
Db 2359 KVDKXTEVPKFTTNDDRRQVKEGEDVKFTANVEGYPEPSVAWTLNGEVPKSHPNITVTDK 2418
QY 776 P-----APTTPKET-----APTTPK----- 790
Db 2419 DGEHTIEISAVTPEQAGELSCAATNPVGSKKRDVQLAVKKVGDAPTAKNLEDRLLITEGE 2478
QY 791 ----- 790

Db 2479 LTLMDAKLNTVKKPKKITWLKOGVEITSDGHYKIVBEEDGSLKLSILOTKLEDKGRITIK 2538
QY 791 -----EPAPTT----- 796
Db 2539 AESEFVAGCSASLVGVKGRPAKPAFQSDIAFINLUTEGDTLECKLLITGDPFFVKWVI 2598
QY 797 -----PKK-----PART 803
Db 2599 GTQLVCATEDEISNANGVYTMKIHGVTADMTGKIKCVAYNKAGEVSTEGPKVWAPIV 2658
QY 804 TPET-----PPPTT-----SEVSTPTTTKE 823
Db 2659 EFETSLCDATCREGDTLKLRPAVLGEPFVSVWYVNGKLEBSQNIKIHSEKGTIVTIIK 2718
QY 824 PTT-----IHK----- 829
Db 2719 DIICDYGQVVCEAINIYKATSEATLLVLRGEPDPFLEWLSNVRRARTGTVVHKVFT 2778
QY 830 -----SPD----- 832
Db 2779 GDFKPSLTWYNNKEILNSDLYTIVTDDKTSLTILNSFNPDPVHVGEIICAEADAGEVSC 2838
QY 833 -----ESTPELSAE-----PTPKALENSPK----- 852
Db 2839 TANNITVTSDFMSESESAQAEBFVGDDLTEDESLREEMHRTPTPW--APKFITKMD 2895
QY 853 -----EPGVPTTK----- 860
Db 2896 TKAKGHSVFCVCPDPTKGVCCKWLKOGKBIELIARIVQTRTGPBGHITOELVLDNVT 2955
QY 861 -----TPAAT-----KPEMTTAKDKTTERDLRT-- 885
Db 2956 PEDAGKYTCIVENTAGDKICEATLVIESLEKSEKKAPEFIVALQDKTKTSEKVLBC 3015
QY 886 -----PETTAAPKPKMTKETATTTBKT--TESKITATTTQ-----VTSTTQDTTTPFKITT 933
Db 3016 KVIGBPKPKVSWLHDNVSEKPNSEKTTIQESITVESVEGVERVITTSSE----- 3066
QY 934 L-----RTTLTAPKVTITTKT-----TITTEIMNKP----- 959
Db 3067 LSHQGYTCIAENTEGTSKTEAPLTVQGEAPVTKELQNKELSIGEKLVLSGVSKSQP 3126
QY 960 -----ETAPKPKDR-----ATNSKAT 976
Db 3127 HVDFYSEFTTKVETKITSSRIAEIHQDQTNTHWRMVISQITKEDIVSYKAIATNSIGTA 3186
QY 977 PKQKPT--KAP-----KKPTSTK-----KP----- 995
Db 3187 TSTSKITTKVEAPVFPQGLUKTSVKEKEIKMVKVGSAPDVENFKDKPVSSEGNHEM 3246
QY 996 KTMPR-----VRPKPTT-----PT-PRKMTSTM 1017
Db 3247 KKNPETGVFTLVVQAAATTDAGKYTAKSNAPGATSAESSAAEVTQSLEKPTFVRELVTTE 3306
QY 1018 PELNPTSRIAEAMLOTTTR-----PQTPNSKLV----- 1046
Db 3307 VKINET-----ATLSVTVTKGVDPDPVSEWLVKQGPVQDSSHVIAKVEGSGSYITIKDAR 3361
QY 1047 -----EYNPKSESDAGGAGETPHMLLR---PHVFMPEVTP 1078
Db 3362 LEDSGKYACRATNP---AGEAKTEANFAVVKLVPPPEVEKLS 3402

RESULT 12

T25697
hypoetical protein F16f9.2 - Caenorhabditis elegans
C;Species: Caenorhabditis elegans
C;Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 09-Jul-2004
C;Accession: T25697
R;Fulton, B.
submitted to the EMBL Data Library, August 1996
A;Description: The sequence of C. elegans cosmid F16f9.
A;Reference number: Z20071


```

652  ---ITTE-----GSTTTEPTTTAIFAEASTGIITDEETSTSTTPEITSTKE-- 698
733  PTPKGPAPTTLKEPAPTT-----KKPAPKELAPTTTGP- 768
699  IVTESAITQSVSVVSSITPQLPWRKAIWNFKHNLEVLKCKRLLKKESTSTGSD 758
769  -----TSTGSKPAPTPPKETASTTPKEPAPTTPKKPAPTPE--TPEPT 811
759  SSETTTVAENIDEVITTEKVVQVPTPTTSTXSTQETTTTTTTTTTEKTSKTTBKPT 818

```

[illegible]

| | | | |
|----|-----|--|-----|
| Qy | 871 | TTAKDKTTERDLRTTETTTA----- | 891 |
| Dh | 852 | TAATNTNTTSP-----TSPDTTSSAAETTCESDENTAT | 908 |

QY 892 APKMTKEATTEKTTBESKIATTTQ-VTST---TTQDTTFPKITTLKTTTAPKVT--- 944

Db 909 VPKKHITVVKPAETTSAAASITTTTETPTTETKSTLTLEPIEATLTNEVTGPAFVTCAP 968

| | | | |
|----|-----|--|------|
| QY | 945 | TTKTKTITTTIMNKPBTAKPKDRATNSKATTPKQ-----KPTKAPK | 987 |
| Dh | 969 | VDEFTTNTTITLISK-----TNNQTSIQKPTTISKDTSISLISGCSFTKAPM | 1018 |

QY 988 KPTSTKKPKTMRVRVRKPKTTPTRKMTSTMPELN-----PTSRFAEAMLQTTT 10355

```

Db
1019 APTI-----HTTIDAAFVIAATEASLNDGSKXIIDEAQPTDERRA-----1059

```

| | | | | | |
|----|------|-----------------|-------|-------------------------------------|------|
| Qy | 1036 | RPNQTENSKLVEVNP | KSEDA | GAGAEGETPHMLLRPHVFMEVTPDMDYLPVPNQGI | 1091 |
| Dp | 1060 | --OPTN-- | ---- | EVDKMFPEKRTREORT | 1080 |

[illegible]

RESUL 13
T18535
high molecular mass nuclear antigen - chicken (fragment)

C:Species: Gallus gallus (chicken)
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 09-Jul-2004
C:Accession: M18535

R. Shimada, K.; Harata, M.; Mizuno, S.
J. Cell Sci. 110, 3031-3041, 1997

A1:Title: A nuclear matrix-associated high molecular mass nuclear antigen, HNNA
A1:Reference number: Z18955; MUID:9803440; PMID:9365273
A1:Accession: T18535

A_Status: preliminary; translated from GB/EMBL/DDBJ
A_Molecule type: mRNA
A_Residues: 1-1153 (SH1)

A; Cross-references: UNIPROT:O57580; EMBL:D88440; NID:d117138; PID:d1025045; P

Query Match 14.6%; Score 861.9; DB 2; Length 1151;
Best Local Similarity 22.3%; Pred. No. 4.3e-10;
Matches 292; Conservative 92; Mismatches 353;
Indels 571; Caps 56;

Qy 110 PTSPSSK-----KAPPPSGASQTIKSTTKRSPKPPNKKTKK 147
|||.||| ||| ||| ||| ||| |||

| | | | | |
|------------|----|---|-------------|----|
| D b | 27 | P T A E P R K W V I A E L H F A A P Q D P P K W V I G G A P P P G T | E P T P P S | 70 |
|------------|----|---|-------------|----|

| | | |
|----|---|-----|
| QY | VIESEELLEIKVORDNKKNRKTKPT-----PKPPVVDEAGSGGLGNG | 187 |
| Db | 71 -----XPTDGDADAAPKASAELETSPPPPAGESPPDGKAPSGA--G | 107 |

```

QY      188  DFKVTTPTDSTTQHNNKVSTSPKITTAKPINPRESLPPNSDTS-----229

```

Db 108 EAEAGTTPPSQ-----GPACTPPSQGAAGAPKGDGTAQPSGTKSGA 149

```

QY      : | | | | | : | | | | | : | | | | | : | | | | | : | | | | | :
230  --EISLNVNKEITVEIK-----ETIINKQISIDGKEKTIISAKETIQSLEK----- 213

Dc 150 DGKFAAQDVEKATTAAATEARPAASAASPTVPKATA---EATV7TAASQSAPKAATDAAV 205

```

Qy 274 TSAKDLPATSKVLAKP-----TPKAETTKGSPALTPKEP----- 308

— — — — —


```
QY 345 KSAPPTPKKPAPT-----TTKEPAPTTPKEPAPTTPKKEPA 379
Db 971 KRPPQTPKEAQALEDLTGFKELFQTPGHTTEAVAAGKTTKMPCESSPPESADT----- 1024
QY 380 PTTTKSAPTPPK----- 392
Db 1025 PTSTRPQKTPLEKRDVQKELSAKLKLTQTSGETTHDQVGGEDKSNAPRETAKQKLD 1084
QY 393 -PAPTTPKPAPTTPKEPA-----PTTPK-----BPTPT- 420
Db 1085 PAASVTGSGRHPT-KEKAQPLEDLAGWKELFQTPVCTDKPTTHEKTKIACRSQDPVD 1143
QY 421 TPKEPAPTTKE-----PAPT----- 435
Db 1144 TPTSSKPSQKSLRKVDVEEFPALRKRTPSAGKAMHTPKPAVSGEKNIYAFMGTPVQKL 1203
QY 436 -----TPKEPA-----PTA 444
Db 1204 DLTENLTGSKRLQTPKEKAQALEDLAGFKELFQTRGHTTEBMTNDTKAVACKSQDPL 1263
QY 445 PKKEPA-----PTTPKEPAPT----- 459
Db 1264 DRNPASSKRLKTSGLKGVGBELLAVGKLQTSGETTHTEPTGDGSKMAKPMESPKQ 1323
QY 460 -----TPKEPA-----PTTPKEPAPT-----E 479
Db 1324 ILDSAASLTGSKRLQTPKGEVPEDLAGFIELFQTPSHTE-SVTNEKTTKVSYRASQ 1382
QY 480 P-----APTTPKSAP----- 489
Db 1383 PDLVDPTSSKQPKESLRKADTEEFLEAFKQTPSAGKAMHTPKPAVGEKNDINTFLT 1442
QY 490 -----TTTKEPA-----PT-----TTK-----SA 503
Db 1443 PVQKLDQGNLPGSNRRLOTRKEKAQALBELTGFRLEFQTPCTDNPDADEKTTKILCKS 1502
QY 504 PTTTPKEPSTTKE-----PA----- 519
Db 1503 PQSDPADTPNTKQPKRSLRKADVEEFLAFKRLTPSAGKAMHTPKPAVGEKNDINTFV 1562
QY 520 -----PTTPKEPAPT-----TP----- 531
Db 1563 GTPVEKLDLLGNLPGSKRRPQTPKEKAQALEDLAGFKELFQTPGHTTEBMTDDKITEVSC 1622
QY 532 KKPAPTTPKEPA----- 543
Db 1623 KSPQDPVKTPTSSKORLISLKGVGKBEVLPVQKLQTSKTTQTHRETADGKSIAK 1682
QY 544 -----PTTPKEPA-----PTTPKPAPT 561
Db 1683 FKESAKQMLDPANYGTGMERNWPTPKAEQSLLEDLAGFKELFQTPDHTEESTDDKTKI 1742
QY 562 APKEPAP-----TTPKET--APTTP-----KKLTPPT----- 586
Db 1743 ACKSPPPESMDPTSTRRRPKTPGLGRDIVEELSALKQLTQTHDQVGGEDKGINVFR 1802
QY 587 ---PEKLAP---TTPKEPAPTTP-----EELA-----PTTPEEPTPTTP 619
Db 1803 ETAKQKLDPAASVTGSKRQRPKPGKAQPLEDLAGLKLQTPVCTDKPTTHEKTKIAC 1862
QY 620 EEPAP-----TTPKEPA-----TTPKAAA----- 631
Db 1863 RSPQDPVGTPTTFKPSKSLRKADVEEELSALRKRTPSVGKAMDTPKPAAGDEKDKMA 1922
QY 632 -----PNTTPKEPA-----PTTPKEPAPT 650
Db 1923 FMGTPVQKLDLPONLPGSKRWPTQTPKEKAQALEDLAGFKELFQTPQTDKPTDEKTKIA 1982
QY 651 PKEPAP-----TTPKET-----APTTPK----- 668
Db 1983 CKSPQDPVDTPASTKQPKENLRKADVEEFLALRKRTPSAGKAMDTPKPAVSEKIN 2042
QY 669 -----GTAPTTLKEPAPTTPKKPAP-----KELAPTTPKPTSTSDK- 706
```

```
Db 2043 TFFVETPVQKLDLLGNLPGSKRQ--PQTPKEKAQALEDLVGFKELFQTPGHTTEBMTDDKI 2100
QY 707 -----PAP-----TTPKGTAPT----- 718
Db 2101 TEVSKSPQSPESFKTSRSSKQRLKIPLVKVDKKEEPLAVSKLRTSGETTQTHTEPTGDS 2160
QY 719 -----TPK-----TPKEPA----- 732
Db 2161 KSIKAFKESPAQIILDPAAASVTGSRQLRTRKAKARALEDLVDFKELFSAFGHTTEBMTID 2220
QY 733 -----PTTPKGTAPTTLKEPAPTTPKKPAPKELAP-----TTTKGTSTITSKPA- 777
Db 2221 KNTKIPKSPPELTDATSTKRCP-KTRPRKEVKEELSAVERLTQTSGETTHTEHPKAS 2279
QY 778 -----PTTPKE-----TA 785
Db 2280 GDEGKVLKQRAKKPNPVEEPPSRPRAPKPKAQPLEDLAGFTELSETSCHTQBSLTA 2339
QY 786 PTTTPKEPAPTTPKPAPTTPETPTTSEVSTPTTTPKEPTTIHKSPDESTEPELSAETTP- 844
Db 2340 GKATKIPCESPPLVWDTTASTKRLRTRVQKVQKEEPSAV-KFTQTSGETTHDADKEPA 2398
QY 845 -----KALENSPKPEGPVPT-----KTBA--TKPEMTT 871
Db 2399 GEDGKIKALKESAKQTPAPAAASVTGSRRRPRAPRESAQAIEDLAGFKDPAGHTEESMT- 2457
QY 872 TAKDTTERDLRTPE--TTTAAFPKMTKETATTTEXTITE--SKITATTQVISTTTQ 924
Db 2458 --DDKTKIPCKSSPELEDATSSKRRPRTRAQKVEVKEELLAVGKLQTSGETTHDKE 2515
QY 925 -----DTTPFKITTLKTTLAPKVTITTKITITTEIMNKPEETAKKDR- 969
Db 2516 PVGEGKGTAFK-----QPAKRNVDADVIGSRQRPAPKKAQPLEDLASF 2562
QY 970 -----TNSKATTPKQPKTK-----APK-KPT----- 990
Db 2563 QELSQTPEHTEELANGAADSFSAKQTPDSGKELKISRRLVIRAPKVPVGDVSTRDPV 2622
QY 991 -----STYKPKTMPRVRKPK 1005
Db 2623 KSQSKSNTSLPPLPKRGGKQGSVTGKRLRCMPAPEEIVEELPASKKORVAPRAGKS 2682
QY 1006 TTPT--PRKMTSMPELNPISRIAEAMLOTTTPNQ-----TPNSKLVEVNFKSEDAGGA 1058
Db 2683 SEPVVIMKRSURTSAKRIEPAEELNSNDKINKKEHLQDSVPENKGISLRSRQDKTEA 2742
QY 1059 EGETHMLL-----RPHVFMPEVTPDMXYLPRVFNQGIINP 1095
Db 2743 BQQITEVFVLAERIEINRNEKKP---MKTSPEND-----IQNP 2777
```

RESULT 15

A48666
cell proliferation antigen Ki-67, long form - human
C:Species: Homo sapiens (man)
C:Date: 03-May-1994 #sequence_revision 03-May-1994 #text_change 09-Jul-2004
C:Accession: A48666
R:Schluter, C.; Duchrow, M.; Wohlenberg, C.; Becker, M.H.G.; Key, G.; Flad, H.D.; Gerdt
J. Cell Biol. 123, 513-522, 1993
A:Title: The cell proliferation-associated antigen of antibody Ki-67: a very large, ubiq-
ins.
A:Reference number: A48666; MUID:94043435; PMID:8227122
A:Accession: A48666
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-3256 <SCH>
A:Cross-references: UNIPROT:P46013; EMBL:X65550; NID:g415818; PID:CAA46519.1; PID:g4151
C:Keywords: alternative splicing; cell cycle control; nucleus; tandem repeat
P:29-91/Domain: kinase interaction domain homology <KIH>

Query Match 14.5%; Score 855.1; DB 2; Length 3256;
Best Local Similarity 15.9%; Pred. No. 2.5e-09;

Matches 418; Conservative 132; Mismatches 450; Indels 1629; Gaps 100;

```

QY 4 KTLPIYLLLLSVFVIQVSSQDLSSCAGRCGEYSRDATC----- 44
DB 600 KTAPE-----ASSSKSQTEVPKRGGE--RVATCLQKRVSISRGHDILQWIC 643
QY 45 -----N 45
DB 644 SKRRSGASEANLIIVAKSWADVVVLGAKQTQTKVIKHGQPSNMNKKQRRPATPKPVGVEH 703
QY 46 CDYNCQHYMECCP-----DFKR----- 62
DB 704 SQFSTGHANSPTIIIGKAHTEKHVHPARPYRVLANFISNQKMDFKEDLSGIAEMFKTPV 763
QY 63 -----VCTAELS-----CKGRCFESFERGECDDAQCCKYKCCPDYFSFCAEV--- 107
DB 764 KEQPQLTSTCHIAISENELLGQFGTDSGEE-----PLLPTSESFGNVPFS 812
QY 108 -HNPTSPPSSK-KAPPP-----SGASQTIKST 132
DB 813 AQNAKQPSDKCSASPPLRQRCIRENGVAKTPRNTYKMTSLETKTSDTETEPSKTV-ST 871
QY 133 TKRS-----PKPPNKKKT----- 145
DB 872 VNRSGRSTEFRIQKLPVBSKSEETNTEIVECILKRGQKATILLOQRREGEKMEIERPPET 931
QY 146 -KKVIESEETEKVDONKNRT-KKPTPKPPVDEAGSLDNGDFKVTTPDT----- 196
DB 932 YKENIELKENDENKAMKASRTWGOKCAPMSDLTD-----LKSLPDTELMKQTA 980
QY 197 -----STTOHNKSVSPKITTAKINPRPSLPP-----NSDTS-----KETS 233
DB 981 RQONLLQTDHAKAPKSEKGIK-KPCQSLOPEPINTPTHTKQOLKASLGKVGKVEEL 1038
QY 234 LVNKEITVETKETTNTKQOTSDQKE-----KTSAKETO 269
DB 1039 LAVGKFRTP-SGETHTHREBPAGDGSIRTFKESPKQLDPAARVTGMKMPRTPKBEAQ 1097
QY 270 SIEKTSAKDLAPTSKVLAKPTPKAE-----TTTK----- 298
DB 1098 SLE-----DLAGFKELFQTPGSEBSMTDEKTKIACKSPPPESVDTPSTIKQWPKRSLR 1152
QY 299 -----GPALTTPKEP-----TPT----- 311
DB 1153 KADVEEFELALRKLTPSAGKAMLTTPKAGGDEKOIKAFMGTPVQKLDLAGLPGSKRLQ 1212
QY 312 TPKEPAST-----TPKEP-----TPPTIKSAPT-- 334
DB 1213 TPKEAQALEDLAGFKELFQTPGHTBELVAAGKTTKIPCDSPQSDPVDTPSTIKQPKRS 1272
QY 335 -----TPKEPAP-----TTT 344
DB 1273 IRKADVEGELLACRLNMPKAGKAMHTPK-----PSVGEEDKIIIFVGTVPQKLDLTENLTGS 1329
QY 345 KSAPTTPKEPAPT-----TTKEPAPTTPKEPAPTTPKEPA 379
DB 1330 KERQTPKEEAQALEDLGTGFKELFQTPGHTBEAAGKTTKMPCESSPPESADT----- 1383
QY 380 PTTTKSAPTTPKE----- 392
DB 1384 PTSTRQPKTPLEKRDVQKELSALKKLTQTSGETHTTKVPGEDKSNAPRETAQKLD 1443
QY 393 -PAPTTPKPAAPTTPKEPA-----PTTPK-----EPTPT- 420
DB 1444 PAASVTGSKRHPKT-KKAQAQLEDLAGWKELFQTPVCTDKXTTHEKTKTKIACRSQDPVD 1502
QY 421 TPKEPAPTTPKE-----PAPT----- 435
DB 1503 TPTSSKQPSKSLRKVDVEEFELALRKLTPSAGKAMHTPKPAVSGEKNIYAFMGTVPQKL 1562
QY 436 -----TPKEPA-----PTA 444
DB 1563 DLTENLTGSKRLQTPKEKAQALEDLAGFKELFQTPRGHTESMTNDKTAACKVACKSQPDL 1622

```

```

QY 445 PKKPA-----PTTPKEPAPT----- 459
DB 1623 DKNPASSKRLKTSLGKVGKVELLAVGLTQTSGETHTHTPTGDKGSKMAFMSPKQ 1682
QY 460 -----TPKEPA-----PTTKPSPTTPK-----E 479
DB 1683 ILDSAASLGSKQKLTGPKSEVPEDLAGFIELFQTPSHTK-SMTNEKTKVSVRASQ 1741
QY 480 P-----APTTPKSAP----- 489
DB 1742 PDLVDTPTSSKQPKRSLRKADETEEFALPRKQTPSAGKAMHTPKPAVGEEDKINTFLGT 1801
QY 490 -----TTTKEPA-----PT-----TTK-----SA 503
DB 1802 PVOKLDPQNLPGSNRRRLQTRKEKAQALBELTGFRELFPQTPCTDNPDADEKTKKILCKS 1861
QY 504 PTTTPKEPSPTTPKE-----PA----- 519
DB 1862 PQSDPADTPTNTKQPKRSILKKADVEEFELAPRKLTPSAGKAMHTPKAAVGEEDKINTEV 1921
QY 520 -----PTTPKEPAPT-----TP----- 531
DB 1922 GTPVEKLDLGNLPGSKRRFPQTPKEKAKALEDLAGFKELFQTPGHTESMTDDKITEVSC 1981
QY 532 KKPAPTTPKEPA----- 543
DB 1982 KSPQDPVKPTPTSSKQRLKISLGKVGKVEEVLVPGKLTQTSKGTQTHRETAGDGSKIA 2041
QY 544 -----PTTPKEPA-----PTTKKPAPT 561
DB 2042 FKESAKQMLDPANYGTMERWPRTPKEEAQSLLEDLAGFKELFQTPDHTDEESTDDKTKI 2101
QY 562 APKEPAP--TPPKET--APTTP-----KLTPTT----- 586
DB 2102 ACKSPPEMDPTSTRRPKTPGLKRDIVBELSALKQLTQTHTDKVPCEDEKGINVFR 2161
QY 587 ---PEKLAP---TTPEKPAPTTP-----EELA-----PTTPEPTPTTP 619
DB 2162 ETAKQLDPAASVTGSKRQRPRTPKGAQPLEDLAGLKFQTPVCTDKPTTHEKTKIAC 2221
QY 620 EEPAP-----TTPKAA----- 631
DB 2222 RSPQDPVGTPTIFKQSKRSRKADVBEESALARKRTPSVGKAMDTPKPAGDEKDKMA 2281
QY 632 -----PNTPKPEPA-----PTTPKEPAPT 650
DB 2282 FMGTPVQKLDLPGNLPGSKRWQTPPKKAQALEDLAGFKELFQTPGTDKPTTDEKTKIA 2341
QY 651 PKEPAP--TTPKET-----APTTPK----- 668
DB 2342 CKSPQDPVDTPASTQRPKRNLRKADVBEELALRKLTPSAGKAMDTPKPAVSBDEKNIN 2401
QY 669 -----GTAPTTLKSPAPTTPKPPAP-----KELAPTTTKPTSTTSK- 706
DB 2402 TFVETPVQKLDLGNLPGSKRQ--FOTPKKABELEDLVGFKELFQTPGHTESMTDDKI 2459
QY 707 -----PAP-----PTPKGTAPT----- 718
DB 2460 TEVSCKSPQESFKTSRSSKQRLKIPLVKVDMKEPPLAVSKLRTSGETTQTHTEPTGDS 2519
QY 719 -----TPK-----TPKEPA----- 732
DB 2520 KSIKAPKESPKQLDPAASVTGSRQRLRTRKKAQALEDLVDFKELFSPAGHTESMTID 2579
QY 733 -----PTPKGTAPTTLKSPAPTTPKPPAPKELAP-----TTTKGPTSTTSKPA- 777
DB 2580 KNTKIPCKSPPELDTATSTKRCP-KTRPRKEVKEELSAVERLTQTSQSTHTEKPAAS 2638
QY 778 -----PTTPKE-----PTA 785
DB 2639 GDEGIKVLQRAKKKNPVVEEPPSRRRPRAPKKAQPLEDLAGFTELSETSGHTQESLTA 2698

```

```
QY 786 PTPKPAATTPKPAATTPPTTSEVSTPTTKEPTTIHKSPDESTPESAEPTP- 844
Db 2699 GKATKIPCESPLEVVDITASTKRHLTRVQVQVKEPSAV-KFTQTSGETTDADKEPA 2757
QY 845 -----KALENSPKERGVPIT-----KTPAA--TKPEMTT 871
Db 2758 GEDKGKALKESAKQTPAASVTSRRPRAPRESAQAIEDLAGFKDPAAGHTEESNT- 2816
QY 872 TAKDKTTERDLRTPE---TTTAAAPKMTKETATTEKTE---SKITATTQVTSITQ 924
Db 2817 --DDKTKTIPCKSSPELEDATSSKRPRTRAQKVEVKEELLAVGKLQTSGETTHDKE 2874
QY 925 -----DTTPEKITTILKTTLAPKVTTTKTITTEIMNKPEETAAPKORA----- 969
Db 2875 PVGEGKGTKAFK-----OPAKNVDAEDVIGSRQPRAPKEKAQPLEDLASF 2921
QY 970 -----TNSKATTPKPKTK-----APK-KPT----- 990
Db 2922 QELSQTPGHTBELANGAADSFTSAPKQTPDSGKPLKISRVLRAPKVEPVGDVVSTRDPV 2981
QY 991 -----STKKPKTMPVRKPK 1005
Db 2982 KQSKSNTSLPLPFKRGKGDSVGTGKRLKCMFAPEIIVEELPASKKQKVAPRARGKS 3041
QY 1006 TTPT---PRKMTSTWPELNPTSRIAEAMLQTTTRNQ---TPNSKLVENVNPKSEDAGGA 1058
Db 3042 SEPVIKESLRTSAKRIPAEELNSNDMKNKEHKLODSVPEKNGISLSRRQDKTEA 3101
QY 1059 EGETHMLL-----RPHVFMPEVTPDMVYLPRVPNQGIINP 1095
Db 3102 EQQITEVFVLABRIEINRNEKFP---MKTSPEMD-----IQNP 3136
```

Search completed: October 13, 2004, 11:58:11
Job time : 68.3652 secs

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: October 13, 2004, 11:23:49 ; Search time 84.5733 Seconds
(without alignments)
4449.477 Million cell updates/sec

Title: SEQ1-D
Perfect score: 576
Sequence: 1 MAWKLPYIYLLLSVFVIQ.....DMDYLPVFNQGIINPMLS 1049

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.1

Searched: 2002273 seqs, 358729299 residues

Total number of hits satisfying chosen parameters: 2002273

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A Geneseq 23Sep04: *
1: Geneseqp1980s: *
2: Geneseqp1980s: *
3: Geneseqp2000s: *
4: Geneseqp2001s: *
5: Geneseqp2002s: *
6: Geneseqp2003as: *
7: Geneseqp2003bs: *
8: Geneseqp2004s: *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Query % | Score | Match | Length | DB ID | Description |
|------------|---------|-------|-------|--------|----------|--------------------|
| 1 | 5546.9 | 99.5 | 1299 | 4 | AAm24322 | AAm24322 Human EST |
| 2 | 5546.9 | 99.5 | 1404 | 2 | AAr26049 | AAr26049 MSF preu |
| 3 | 5546.9 | 99.5 | 1404 | 4 | AAr29773 | AAr29773 Human meg |
| 4 | 5546.9 | 99.5 | 1404 | 4 | AAr60568 | AAr60568 Human meg |
| 5 | 5546.9 | 99.5 | 1404 | 8 | ADm98014 | ADm98014 Human meg |
| 6 | 5546.9 | 99.5 | 1415 | 4 | AAU32262 | AAU32262 Novel hum |
| 7 | 5523.9 | 99.1 | 1404 | 7 | ADK65839 | ADK65839 Angiogene |
| 8 | 5384.3 | 96.6 | 1320 | 7 | ADK65819 | ADK65819 Angiogene |
| 9 | 5358.7 | 96.1 | 1270 | 8 | ADK67911 | ADK67911 Human ext |
| 10 | 5344.6 | 95.9 | 1311 | 8 | ADK67912 | ADK67912 Human ext |
| 11 | 2920 | 52.4 | 546 | 4 | ABU53252 | ABU53252 Human tes |
| 12 | 2888 | 51.8 | 551 | 4 | ABU53253 | ABU53253 Human tes |
| 13 | 2880.1 | 51.7 | 902 | 4 | ABU53254 | ABU53254 Human tes |
| 14 | 2757 | 49.4 | 513 | 4 | ABU53254 | ABU53254 Human tes |
| 15 | 1730.1 | 31.0 | 452 | 5 | AAr80041 | AAr80041 Human meg |
| 16 | 1409.9 | 25.3 | 538 | 5 | AAr80041 | AAr80041 Human meg |
| 17 | 1346.8 | 24.2 | 5179 | 6 | AAm24516 | AAm24516 C899P pre |
| 18 | 1346.8 | 24.2 | 5179 | 6 | ABP55365 | ABP55365 Human col |
| 19 | 1346.8 | 24.2 | 5179 | 6 | ABO07258 | ABO07258 Human p33 |
| 20 | 1346.8 | 24.2 | 5179 | 7 | ADd48091 | ADd48091 Human pro |
| 21 | 1346.8 | 24.2 | 5179 | 7 | ADd44998 | ADd44998 Human pro |
| 22 | 1174.4 | 21.1 | 232 | 5 | AAU11261 | AAU11261 Human HAP |
| 23 | 1128.7 | 20.2 | 1664 | 4 | AAW43106 | AAW43106 C. thermo |
| 24 | 1099.6 | 19.7 | 717 | 4 | ABU53144 | ABU53144 Human tes |
| 25 | 1097 | 19.7 | 214 | 4 | ABU53255 | ABU53255 Human tes |

| | | | | | | |
|----|--------|------|------|---|----------|--------------------|
| 26 | 1088.5 | 19.5 | 8991 | 6 | ABU08487 | ABU08487 S. pneumo |
| 27 | 1068 | 19.2 | 1795 | 4 | ABB69806 | ABB69806 Drosophil |
| 28 | 1042.8 | 18.7 | 763 | 3 | AAG38942 | AAG38942 Arabidops |
| 29 | 1039.7 | 18.6 | 1049 | 4 | ABB61364 | ABB61364 Drosophil |
| 30 | 1020.5 | 18.3 | 2284 | 4 | ABB71434 | ABB71434 Drosophil |
| 31 | 987.9 | 17.7 | 770 | 4 | ABU53141 | ABU53141 Human tes |
| 32 | 981 | 17.6 | 778 | 4 | ABU53143 | ABU53143 Human tes |
| 33 | 978.7 | 17.6 | 5703 | 8 | ADL23265 | ADL23265 Human MUC |
| 34 | 975.9 | 17.5 | 2112 | 4 | ABB60403 | ABB60403 Drosophil |
| 35 | 974.4 | 17.5 | 692 | 4 | ABU53155 | ABU53155 Human tes |
| 36 | 964.3 | 17.3 | 4315 | 5 | ABP43908 | ABP43908 MUC5B par |
| 37 | 964.1 | 17.3 | 717 | 4 | ABU53150 | ABU53150 Human tes |
| 38 | 964.1 | 17.3 | 717 | 4 | ABU53149 | ABU53149 Human tes |
| 39 | 964.1 | 17.3 | 717 | 4 | ABU53151 | ABU53151 Human tes |
| 40 | 964.1 | 17.3 | 717 | 4 | ABU53145 | ABU53145 Human tes |
| 41 | 964.1 | 17.3 | 717 | 4 | ABU53148 | ABU53148 Human tes |
| 42 | 964.1 | 17.3 | 717 | 4 | ABU53147 | ABU53147 Human tes |
| 43 | 964.1 | 17.3 | 717 | 4 | ABU53146 | ABU53146 Human tes |
| 44 | 952.1 | 17.1 | 745 | 4 | ABU53142 | ABU53142 Human tes |
| 45 | 951.6 | 17.1 | 1538 | 7 | ADI21202 | ADI21202 Novel hum |

ALIGNMENTS

RESULT 1

AAW24322
ID AAW24322 standard; protein; 1299 AA.

AC AAW24322;
DT 12-OCT-2001 (first entry)
DE Human EST encoded protein SEQ ID NO: 1847.

Human; sheep; pig; cow; fruit fly; yeast; hamster; macaque; horse;
tomato; monkey; dog; sea urchin; expressed sequence tag; EST;
diagnostics; forensic test; gene mapping; genetic disorder; biodiversity;
gene therapy; nutrition.

OS Homo sapiens.

PN WO200154477-A2.

PD 02-AUG-2001.

PF 25-JAN-2001; 2001WO-US002687.

PR 25-JAN-2000; 2000US-00491404.

PR 17-JUL-2000; 2000US-00617746.

PR 03-AUG-2000; 2000US-00631451.

PR 15-SEP-2000; 2000US-00663870.

PA (HYSE-) HYSEQ INC.

PI Tang YT, Liu C, Zhou P, Qian XB, Wang Z, Chen R, Asundi V;

PI Cao Y, Drmanac RA, Zhang J, Werhman T;

PI WPI, 2001-476164/51.

PI N-PSDB; AAH98981.

PT Isolated polypeptide for treatment of diseases, diagnostics, raising

PT antibodies and research use.

PS Claim 20; Page 1198-1201; 12/5pp; English.

CC The present invention provides the protein and coding sequences of novel proteins from a variety of organisms, including human, dog cat, horse, cow, pig, hamster, monkey, macaque, yeast, bacteria, fruit fly, sea urchin and tomato. These were derived from expressed sequence tags (ESTs) from the organism of interest. They can be used in diagnostics, forensics, gene mapping, identification of mutations, to assess biodiversity and for nutritional purposes. The present sequence is a

```

CC protein of the invention
XX
SQ Sequence 1299 AA;

Query Match          99.5%; Score 5546.9; DB 4; Length 1299;
Best Local Similarity 92.0%; Pred. No. 8.8e-152;
Matches 1049; Conservative 0; Mismatches 0; Indels 91; Gaps 2;

QY 1 MAWKTLPIYLLLLSVFVQQVSSQ----- 25
Db 1 MAWKTLPIYLLLLSVFVQQVSSQDLSGACGCGEYSDATONCDYNCQHYMECCPDF 60

QY 26 -----ELCKGRCRPFSPRGRECDCAQCKYDKCCPDYSEFCA----- 64
Db 61 KRVCTAELSCKGRCFSPRGRECDCAQCKYDKCCPDYSEFCAEVHNPTSPFSSKKAP 120

QY 65 -----BEHSVSENGQSSSSSSSSSSSSSIW 89
Db 121 PPSGASQTIKSTKSPKPPNKKTKVIESEITEHSVSENGQSSSSSSSSSSSIW 180

QY 90 KIKSSKNSAANRELOKLVKDNKNKRTKKKTPKPPVVDVDEAGSLDNGDFKVTTPDTST 149
Db 181 KIKSSKNSAANRELOKLVKDNKNKRTKKKTPKPPVVDVDEAGSLDNGDFKVTTPDTST 240

QY 150 TQHNKVSTSPKLTAKPINRPSLPNSDTSKETSLSLVNKKETTVEKTTTNNKQTSIDG 209
Db 241 TQHNKVSTSPKLTAKPINRPSLPNSDTSKETSLSLVNKKETTVEKTTTNNKQTSIDG 300

QY 210 KEKTTSAKETQSIKTSAKDLAPTSKVLAKPTPKAETTTKGPALTTPKPTPTPKSPAS 269
Db 301 KEKTTSAKETQSIKTSAKDLAPTSKVLAKPTPKAETTTKGPALTTPKPTPTPKSPAS 360

QY 270 TTPKPTPTTIISAPTTKEPAPTTTKSAPTTKEPAPTTKEPAPTTKEPAPTTKEPAPTTKEP 329
Db 361 TTPKPTPTTIISAPTTKEPAPTTTKSAPTTKEPAPTTKEPAPTTKEPAPTTKEPAPTTKEP 420

QY 330 APTTTKSAPTTKEPAPTTPKKAPTTKEPAPTTKEPAPTTKEPAPTTKEPAPTTKEPAPTTKP 389
Db 421 APTTTKSAPTTKEPAPTTPKKAPTTKEPAPTTKEPAPTTKEPAPTTKEPAPTTKEPAPTTKP 480

QY 390 EPAPTAPKAPPTTKKAPPTTKKAPPTTKKAPPTTKKAPPTTKKAPPTTKKAPPTTKKAPPT 449
Db 481 EPAPTAPKAPPTTKKAPPTTKKAPPTTKKAPPTTKKAPPTTKKAPPTTKKAPPTTKKAPPT 540

QY 450 TTKSAPTTKPSPTTKKAPPTTKKAPPTTKKAPPTTKKAPPTTKKAPPTTKKAPPTTKKAP 509
Db 541 TTKSAPTTKPSPTTKKAPPTTKKAPPTTKKAPPTTKKAPPTTKKAPPTTKKAPPTTKKAP 600

QY 510 APTAPKEPAPTTKETAPTTPKLLPTTPPEKLAPTTPKAPPTTPPELAPTTPEEPTPTT 569
Db 601 APTAPKEPAPTTKETAPTTPKLLPTTPPEKLAPTTPKAPPTTPPELAPTTPEEPTPTT 660

QY 570 PEEPAPTTPKAAPNTKAPPTTKKAPPTTKKAPPTTKKAPPTTKKAPPTTKKAPPTTKKAP 629
Db 661 PEEPAPTTPKAAPNTKAPPTTKKAPPTTKKAPPTTKKAPPTTKKAPPTTKKAPPTTKKAP 720

QY 630 APTTPKPKAPKELAPTTTKETSTTSKAPPTTKGAPTTTKGAPTTTKGAPTTTKGAPTTTKG 689
Db 721 APTTPKPKAPKELAPTTTKETSTTSKAPPTTKGAPTTTKGAPTTTKGAPTTTKGAPTTTKG 780

QY 690 TAPTTLKAPAPTTTKKAPKELAPTTTKGPTSTTSKAPPTTKGAPTTTKGAPTTTKGAPPTTK 749
Db 781 TAPTTLKAPAPTTTKKAPKELAPTTTKGPTSTTSKAPPTTKGAPTTTKGAPTTTKGAPPTTK 840

QY 750 KPAPTTPTPTTSEVSTPTTTKEPTTIHKSPDSETPELSAEPTPKALENSPKPEGVPT 809
Db 841 KPAPTTPTPTTSEVSTPTTTKEPTTIHKSPDSETPELSAEPTPKALENSPKPEGVPT 900

QY 810 TKTPAATKPEMTTAKDKITERRDLTTPPETTTAAPKMTKETATTTKTTESKITATTQV 869
Db 901 TKTPAATKPEMTTAKDKITERRDLTTPPETTTAAPKMTKETATTTKTTESKITATTQV 960

QY 870 TSTTTQDTPPKITTLTKTLTAPKVTTTKTITTEIMNKPBEETAKPKDRATNSKATTPK 929

```


The invention relates to a method of treating osteoarthritis via the administration of a composition comprising the camptodactyl-arthropathy-coxa vara-pericarditis (CACP) protein, or portions of the CACP protein. The composition may further comprise a local anaesthetic. The composition of the invention may be administered via intra-articular or intravenous injection. The human CACP protein is identified in the invention as being megakaryocyte stimulating factor (MSF). The gene encoding CACP protein (MSF) is located on chromosome 1q25-31, and mutations in this gene are responsible for the heritable disorder camptodactyl-arthropathy-coxa vara-pericarditis, in which patients have synovial hyperplasia without evidence of inflammation. CACP protein (MSF) acts as a synovium lubricant, and can be used to lubricate tissue and joints in the treatment of osteoarthritis. The composition may be applied to reduce the symptoms of osteoarthritis (e.g., joint pain, loss of range of movement or joint damage). The present sequence represents human megakaryocyte stimulating factor (MSF, CACP protein). Note: This sequence is not given in its entirety in figure 4 of the specification, although a GenBank accession number was given. This sequence was therefore obtained from GenBank (U70316)

XX SQ Sequence 1404 AA;

Query Match 99.5%; Score 5546.9; DB 4; Length 1404;
Best Local Similarity 92.0%; Pred. No. 9.7e-152;
Matches 1049; Conservative 0; Mismatches 0; Indels 91; Gaps 2;

QY 1 MAWKTLPIYLLLSLVFVIOVSSQ----- 25
DB 1 MAWKTLPIYLLLSLVFVIOVSSQ----- 60

QY 26 -----ELSCKGRCPESFERGECDDAOCKKYDKCCPYESFCA----- 64
DB 61 KRVTAEELCKGCFESFERGECDDAOCKKYDKCCPYESFCAEVHNPSPSSKKAP 120

QY 65 -----FEHSVSNQSSSSSSSSSSSSSIW 89
DB 121 PPSGASQTKSTTKRSPKPPNKKKTKVIESEBITEHSVSNQSSSSSSSSSIW 180

QY 90 KIKSSKNSAANRELQKLVKONKNRKKPKPPVVDVDEAGSGLDNGDKVTPDTSF 149
DB 181 KIKSSKNSAANRELQKLVKONKNRKKPKPPVVDVDEAGSGLDNGDKVTPDTSF 240

QY 150 TQHNKYSTSPKIITAKPINRPSLPNSDTSKETSILVNKETTVEKTTTNKQSTDG 209
DB 241 TQHNKYSTSPKIITAKPINRPSLPNSDTSKETSILVNKETTVEKTTTNKQSTDG 300

QY 210 KEKTSKETSQSTKTSKADLAPTSKVLAKPTKAEITTKGPAITTPKPTTKPERAS 269
DB 301 KEKTSKETSQSTKTSKADLAPTSKVLAKPTKAEITTKGPAITTPKPTTKPERAS 360

QY 270 TTPKEPTTIKSAPTTPKEPAPTTTKSAPTTPKEPAPTTTKSAPTTPKEPAPTTTKEP 329
DB 361 TTPKEPTTIKSAPTTPKEPAPTTTKSAPTTPKEPAPTTTKSAPTTPKEPAPTTTKEP 420

QY 330 APTTKSAPTTPKEPAPTTTKSAPTTPKEPAPTTTKSAPTTPKEPAPTTTKSAPTTPK 389
DB 421 APTTKSAPTTPKEPAPTTTKSAPTTPKEPAPTTTKSAPTTPKEPAPTTTKSAPTTPK 480

QY 390 EPAPKPKAPPTTKSAPTTPKEPAPTTTKSAPTTPKEPAPTTTKSAPTTPKEPAPTT 449
DB 481 EPAPKPKAPPTTKSAPTTPKEPAPTTTKSAPTTPKEPAPTTTKSAPTTPKEPAPTT 540

QY 450 TTKSAPTTPKEPAPTTTKSAPTTPKEPAPTTTKSAPTTPKEPAPTTTKSAPTTPK 509
DB 541 TTKSAPTTPKEPAPTTTKSAPTTPKEPAPTTTKSAPTTPKEPAPTTTKSAPTTPK 600

QY 510 APTAPKEPAPTTTKSAPTTPKEPAPTTTKSAPTTPKEPAPTTTKSAPTTPKEPAPTT 569
DB 601 APTAPKEPAPTTTKSAPTTPKEPAPTTTKSAPTTPKEPAPTTTKSAPTTPKEPAPTT 660

QY 570 PEEPAPPTPKAAAPNTPKSAPTTPKEPAPTTTKSAPTTPKEPAPTTTKSAPTTPK 629
DB 661 PEEPAPPTPKAAAPNTPKSAPTTPKEPAPTTTKSAPTTPKEPAPTTTKSAPTTPK 720

QY 630 APTTPKPKAPKELAPTTTKETSTSDKAPPTPKGAPPTPKAPPTPKAPPTPKG 689
DB 721 APTTPKPKAPKELAPTTTKETSTSDKAPPTPKGAPPTPKAPPTPKAPPTPKG 780

QY 690 TAPTTLKEPAPPTPKAPKELAPTTTKGPTSTSDKAPPTPKETAPTTPKAPPTPK 749
DB 781 TAPTTLKEPAPPTPKAPKELAPTTTKGPTSTSDKAPPTPKETAPTTPKAPPTPK 840

QY 750 KPAPPTPETPPPTTSEVSTTTTKETPTTIHKSPDESTPELSAETPKALENSPK 809
DB 841 KPAPPTPETPPPTTSEVSTTTTKETPTTIHKSPDESTPELSAETPKALENSPK 900

QY 810 TKTPAAKPEMTTAKDKTTERDLRTPTTAAAPKMTKATTTTETTSKITATTTQV 869
DB 901 TKTPAAKPEMTTAKDKTTERDLRTPTTAAAPKMTKATTTTETTSKITATTTQV 960

QY 870 TSTTTQDTTPPKIITLLKTTTLAPKVTTKTITTEIMNKPEETAKEKDRATNSKATTPK 929
DB 961 TSTTTQDTTPPKIITLLKTTTLAPKVTTKTITTEIMNKPEETAKEKDRATNSKATTPK 1020

QY 930 PQKPTKAPKPTSTKPKMTMPRVKPKTTTTPRKMSTMPELNPTSTRIAEAMLQTTTRPN 989
DB 1021 PQKPTKAPKPTSTKPKMTMPRVKPKTTTTPRKMSTMPELNPTSTRIAEAMLQTTTRPN 1080

QY 990 QTPNSKLVEVNPKSEDAGGAGETPHMLLRPHVFMPEVTPDMDYLPVNPQGIINPMLS 1049
DB 1081 QTPNSKLVEVNPKSEDAGGAGETPHMLLRPHVFMPEVTPDMDYLPVNPQGIINPMLS 1140

RESULT 5
ADM98014
ID ADM98014 standard; protein; 1404 AA.
XX AC ADM98014;
XX DT 01-JUL-2004 (first entry)
XX DE Human megakaryocyte stimulating factor (MSF).
XX KW lubricating polypeptide; O-linked oligosaccharide; joint lubrication;
KW CAP; camptodactyl-arthropathy pericarditis; osteoarthritis; human;
KW megakaryocyte stimulating factor; MSF.
XX OS Homo sapiens.
XX PN US2004072741-A1.
XX PD 15-APR-2004.
XX PF 02-JUL-2001; 2001US-00897188.
XX PR 23-APR-1999; 99US-00298970.
PR 24-APR-2000; 2000US-00556246.
XX PA (JAYG/) JAY G D.
XX PI Jay GD;
XX DR WPI: 2004-373948/35.
DR N-PSDB; ADM98015.
XX PT New tribonecin polypeptides and polynucleotides for lubricating joints
PT or other tissues to prevent or treat Camptodactyl-arthropathy-
PT pericarditis syndrome or osteoarthritis.
XX PS Claim 1; SEQ ID NO 1; 34pp; English.
XX CC The invention relates to a lubricating polypeptide and at least one O-
CC linked oligosaccharide. The composition and methods are useful for
CC lubricating joints or other tissues to prevent or treat camptodactyl-
CC arthropathy pericarditis (CAP) or osteoarthritis in mammals. The present
CC sequence represents the amino acid sequence of the human megakaryocyte

QY 90 KIKSSKNSAANRELOKLLKVKONKQRTKKXTPKPPVVDVDEAGSLDNGDFKVTTPDTST 149
 DB 181 KIKSSKNSAANRELOKLLKVKONKQRTKKXTPKPPVVDVDEAGSLDNGDFKVTTPDTST 240
 QY 150 TQHNKYSTSPKITTAKPINSRSLPNSDTSKETSITVNKETTIVETKETTINKOTSIDG 209
 DB 241 TQHNKYSTSPKITTAKPINSRSLPNSDTSKETSITVNKETTIVETKETTINKOTSIDG 300
 QY 210 KETTSKETSITVNKETTIVETKETTINKOTSIDG 269
 DB 301 KETTSKETSITVNKETTIVETKETTINKOTSIDG 360
 QY 270 TTPKEPTTTIKSAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEP 329
 DB 361 TTPKEPTTTIKSAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEP 420
 QY 330 APTTTKSAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPK 389
 DB 421 APTTTKSAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPK 480
 QY 390 EPAPTAPKPAPTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPK 449
 DB 481 EPAPTAPKPAPTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPK 540
 QY 450 TTKSAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEP 509
 DB 541 TTKSAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEP 600
 QY 510 APTAPKPAPTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPK 569
 DB 601 APTAPKPAPTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPK 660
 QY 570 PEEPAPTTTPKAAAPNTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEP 629
 DB 661 PEEPAPTTTPKAAAPNTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEP 720
 QY 630 APTTPKPAKPAELAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEP 689
 DB 721 APTTPKPAKPAELAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEP 780
 QY 690 TAPTTTPKEPAPTTTPKPAKPAELAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEP 749
 DB 781 TAPTTTPKEPAPTTTPKPAKPAELAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEP 840
 QY 750 KPAPTTTPPTTSEVSTPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEP 809
 DB 841 KPAPTTTPPTTSEVSTPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEP 900
 QY 810 TKTPAATPEMTTAKDITRDLATTTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPK 869
 DB 901 TKTPAATPEMTTAKDITRDLATTTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPK 960
 QY 870 TSTTTQDTTPKITTLLKTTLLAPKVTITTKTITTEIMNKPEETAKPKDRATNSKATTPK 929
 DB 961 TSTTTQDTTPKITTLLKTTLLAPKVTITTKTITTEIMNKPEETAKPKDRATNSKATTPK 1020
 QY 930 POKPTKAPKPTSTKPKTTPKAPKPTTPKAPKPTTPKAPKPTTPKAPKPTTPKAPKPTTPK 989
 DB 1021 POKPTKAPKPTSTKPKTTPKAPKPTTPKAPKPTTPKAPKPTTPKAPKPTTPKAPKPTTPK 1080
 QY 990 QTPNSKLVNPKSDAGGAGEPHMLLRPHVFWPEVTPMDVYLPVFNQGIINPMLS 1049
 DB 1081 QTPNSKLVNPKSDAGGAGEPHMLLRPHVFWPEVTPMDVYLPVFNQGIINPMLS 1140

RESULT 8

ADK65819

ID ADK65819 standard; protein; 1320 AA.

XX

AC ADK65819;

XX

DT 06-MAY-2004 (first entry)

XX

DE
 XX
 KW
 KW
 KW
 KW
 OS
 XX
 XX
 FN
 XX
 PD
 XX
 PF
 XX
 PR
 PR
 PR
 PR
 XX
 PA
 XX
 PI
 XX
 DR
 DR
 XX
 PT
 PT
 PT
 PT
 XX
 PS
 XX
 CC
 CC
 CC
 CC
 CC
 CC
 CC
 CC
 CC
 CC
 SQ

AnGIOgenesis-differentially expressed protein ANH0316.
 cytostatic; cardiant; vasotropic; antiarteriosclerotic;
 angiogenesis inhibitor; angiogenesis stimulator; angiogenic index;
 gene expression; cancer; coronary artery disease; myocardial ischemia;
 coronary arteriosclerosis; forensic medicine.
 Homo sapiens.
 WO2003066831-A2.
 14-AUG-2003.
 07-FEB-2003; 2003WO-US003848.
 07-FEB-2002; 2002US-00067482.
 10-JUN-2002; 2002US-00164595.
 16-AUG-2002; 2002US-0403649P.
 03-JAN-2003; 2003US-0437746P.
 (ORIG-) ORIGENE TECHNOLOGIES INC.
 Sun Z, Li X, Kovacs KF, Fan W, Jay G;
 WPI; 2003-731502/69.
 N-PSDB; ADK65818.
 Determining the angiogenic index of a tissue or cell sample using
 expression levels of differentially expressed genes, useful for
 diagnosing or treating cancer, coronary artery disease, myocardial
 ischemia and/or arteriosclerosis.
 Claim 23; SEQ ID NO 58; 296pp; English.
 The invention relates to a method of determining the angiogenic index of
 a tissue or cell sample comprising assessing, in a sample, the expression
 levels of one or more differentially-expressed gene from any of 34 DNA
 sequences, given in the specification, where the levels are indicative of
 the angiogenic index. The methods and compositions of the present
 invention are useful for diagnosing, preventing and/or treating cancer,
 coronary artery disease, myocardial ischemia or coronary
 arteriosclerosis. They can also be used in research, drug discovery and
 forensic medicine involving angiogenesis. This sequence corresponds to
 one of the differentially expressed proteins of the invention.

Sequence 1320 AA;

Query Match 96.6%; Score 5384.3; DB 7; Length 1320;

Best Local Similarity 95.9%; Pred. No. 4.4e-147; Mismatches 28; Indels 7; Gaps 1;

Matches 1013; Conservative 8;

QY 1 MAWKTLPIYLLLLSVFVIQQVSSQELSCGRCFESFERGRCDDAOCKYDKCCPDYE 60

DB 1 MAWKTLPIYLLLLSVFVIQQVSSQELSCGRCFESFERGRCDDAOCKYDKCCPDYE 60

QY 61 SFCAEHSVSENOESSSSSSSSSTWIKSSXNSAANRELOKLL-----KVKONK 113

DB 61 SFCAEHSVSENOESSSSSSSSSTWIKSSXNSAANRELOKLL-----KVKONK 120

QY 114 KQRTKKKPTPKPPVVDVDEAGSLDNGDFKVTTPDTSTTQHNKYSTSPKITTAKPINRPSL 173

DB 121 KQRTKKKPTPKPPVVDVDEAGSLDNGDFKVTTPDTSTTQHNKYSTSPKITTAKPINRPSL 180

QY 174 PNSTSKETSITVNKETTIVETKETTINKOTSIDGKSKTSKETSITVNKETTIVETKETTINKOTSIDG 233

DB 181 PNSTSKETSITVNKETTIVETKETTINKOTSIDGKSKTSKETSITVNKETTIVETKETTINKOTSIDG 240

QY 234 SKVLAKPTPKAETTTKGPALTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPK 293

DB 241 SKVLAKPTPKAETTTKGPALTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPK 300

QY 294 TTKSAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPK 353

| | | | |
|--------|---|---|------|
| QY | 301 | TPKEPAPTTTKBPAPTTTPKEPAPTTTKBPAPTTTKSAPTTTPKEPAPTTTKBPAPTTTPKEP | 360 |
| Db | 258 | TPKEPAPTTTKBPAPTTTPKEPAPTTTKBPAPTTTKSAPTTTPKEPAPTTTKBPAPTTTPKEP | 317 |
| QY | 361 | APTTTPKEPPTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEP | 420 |
| Db | 318 | APTTTPKEPPTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEP | 377 |
| QY | 421 | KBPSPPTTPKEPAPTTTKSAPTTTKBPAPTTTKSAPTTTPKEPAPTTTPKEPAPTTTPKEP | 480 |
| Db | 378 | KBPSPPTTPKEPAPTTTKSAPTTTKBPAPTTTKSAPTTTPKEPAPTTTPKEPAPTTTPKEP | 437 |
| QY | 481 | TPKKPAPTTTPKEPAPTTTPKEPAPTTTKBPAPTTTPKEPAPTTTPKEPAPTTTPKEP | 540 |
| Db | 438 | TPKKPAPTTTPKEPAPTTTPKEPAPTTTKBPAPTTTPKEPAPTTTPKEPAPTTTPKEP | 497 |
| QY | 541 | LAPTTPKEPAPTTPEELAPTTPEEPTTPPEBPAPTTPKAAAPNTPKEPAPTTTPKEP | 600 |
| Db | 498 | LAPTTPKEPAPTTPEELAPTTPEEPTTPPEBPAPTTPKAAAPNTPKEPAPTTTPKEP | 557 |
| QY | 601 | TPKEPAPTTPKETAPTTPKGTAPTTLKBPAPTTPKKBPAPKELAPTTTKEPTSTSDKAP | 660 |
| Db | 558 | TPKEPAPTTPKETAPTTPKGTAPTTLKBPAPTTPKKBPAPKELAPTTTKEPTSTSDKAP | 617 |
| QY | 661 | TPPKGTAPTTTPKEPAPTTTPKEPAPTTTPKGATPPTLKBPAPTTTPKBPAPKELAPTTTKGT | 720 |
| Db | 618 | TPPKGTAPTTTPKEPAPTTTPKEPAPTTTPKGATPPTLKBPAPTTTPKBPAPKELAPTTTKGT | 677 |
| QY | 721 | STSDKAPTTPKETAPTTTPKEPAPTTTPKBPAPTTTPETPPTTSVSTPTTKETPTTHK | 780 |
| Db | 678 | STSDKAPTTPKETAPTTTPKEPAPTTTPKBPAPTTTPETPPTTSVSTPTTKETPTTHK | 737 |
| QY | 781 | SPDESTPELSAPTPKALENSPKPGVPTTKTPAATKPEMTTAKDKTTERDLRTPTT | 840 |
| Db | 738 | SPDESTPELSAPTPKALENSPKPGVPTTKTPAATKPEMTTAKDKTTERDLRTPTT | 797 |
| QY | 841 | TAAPKMTKETATTTKTTESKITATTQTSTTTQDTTTFKTLTKTTTLAPKVTTTKT | 900 |
| Db | 798 | TAAPKMTKETATTTKTTESKITATTQTSTTTQDTTTFKTLTKTTTLAPKVTTTKT | 857 |
| QY | 901 | ITTTIMNKPETAKPKORATNSKATTPKPKPTKAPKPKPTSTKPKTMVVRPKTTPT | 960 |
| Db | 858 | ITTTIMNKPETAKPKORATNSKATTPKPKPTKAPKPKPTSTKPKTMVVRPKTTPT | 917 |
| QY | 961 | PRKMTSTMPENLPTSRIAEAMLTTRPNQTPNSKLVNPNKSDAGAGETPHMLLRP | 1020 |
| Db | 918 | PRKMTSTMPENLPTSRIAEAMLTTRPNQTPNSKLVNPNKSDAGAGETPHMLLRP | 977 |
| QY | 1021 | HVFMPEVTPDMDYLPRVFNQGIINPMLS | 1049 |
| Db | 978 | HVFMPEVTPDMDYLPRVFNQGIINPMLS | 1006 |
| RESULT | 10 | | |
| ID | ADK67912 | | |
| XX | ADK67912 standard; protein; 1311 AA. | | |
| AC | ADK67912; | | |
| XX | ADK67912; | | |
| DT | 06-MAY-2004 (first entry) | | |
| DE | Human extracellular messenger (EXMES) polypeptide. | | |
| XX | Human; extracellular messenger; EXMES; respiratory-gen.; antiallergic; | | |
| KW | antiasmatic; antiinflammatory; antidiabetic; neuroprotective; | | |
| KW | muscular-gen.; antiarthritic; osteopathic; hepatotropic; antiparasitic; | | |
| KW | virucide; fungicide; antiparasitic; protozoacide; antihelminthic; | | |
| KW | cytostatic; gene therapy. | | |
| XX | | | |
| OS | Homo sapiens. | | |
| XX | | | |
| FH | Key | | |
| | Location/Qualifiers | | |

| | | | |
|----|--|--------|--|
| FT | Peptide | 1. -30 | |
| FT | /label= Signal_peptide | | |
| FT | /note= "Spans residues 1 to 18, 20, 21, 24, 29 or 30 | | |
| XX | according to identification method" | | |
| FN | W02004013292-A2. | | |
| XX | | | |
| PD | 12-FEB-2004. | | |
| XX | | | |
| FF | 30-JUL-2003; 2003W0-US024084. | | |
| XX | | | |
| PR | 02-AUG-2002; 2002US-0400810P. | | |
| PR | 19-SEP-2002; 2002US-0412197P. | | |
| PR | 04-OCT-2002; 2002US-0416004P. | | |
| PR | 08-NOV-2002; 2002US-0424862P. | | |
| XX | (INCY-) INCYTE CORP. | | |
| FA | | | |
| XX | Elliott VS, Khare R, Tran UK, Swarnakar A, Marquis JP; | | |
| PI | Richardson TW, Emerling BM, Lindquist BA, Chawla NK, Ramkumar J; | | |
| PI | Lee SY; | | |
| XX | | | |
| DR | WPI; 2004-157116/15. | | |
| DR | N-PSDB; ADK67917. | | |
| XX | | | |
| PS | Claim 60; SEQ ID NO 5; 165pp; English. | | |
| XX | | | |
| CC | The present sequence is that of novel human extracellular messenger | | |
| CC | (EXMES) Incyte ID NO: 7513018CD1 polypeptide. The protein shows homology | | |
| CC | to human megakaryocyte stimulating factor. The invention provides EXMES | | |
| CC | polynucleotides and polypeptides, as well as expression vectors, host | | |
| CC | cells, antibodies, agonists and antagonists, and methods for diagnosing, | | |
| CC | treating or preventing disorders associated with aberrant expression of | | |
| CC | EXMES, especially autoimmune and inflammatory disorders, cell | | |
| CC | proliferative disorders and endocrine disorders, e.g. adult respiratory | | |
| CC | distress syndrome, allergy, anaemia, erythroblastosis foetalis, Crohn's | | |
| CC | disease, diabetes mellitus, myasthenia Gravis, osteoarthritis, | | |
| CC | osteoporosis, pancreatitis, psoriasis, viral, bacterial, fungal, | | |
| CC | parasitic, protozoal or helminthic infections, cancers, autoimmune | | |
| CC | thyroiditis, cretinism, Plummer's disease or thyroid carcinoma. | | |
| CC | Embodiments also provide methods for using the purified EXMES and/or | | |
| CC | their encoding polynucleotides for facilitating the drug discovery | | |
| CC | process, including determining of efficacy, dosage, toxicity and | | |
| CC | pharmacology, and for investigating the pathogenesis of diseases and | | |
| CC | medical conditions. | | |
| XX | Sequence 1311 AA; | | |

| | | | | | | | |
|-----------------------|--------|---|-----------|------------|----|--------|-------|
| Query Match | 95.9%; | Score | 5344.6; | DB | 8; | Length | 1311; |
| Best Local Similarity | 92.3%; | Pred. No. | 6.1e-146; | | | | |
| Matches | 1006; | Conservative | 0; | Mismatches | 0; | Indels | 84; |
| Gaps | 2; | | | | | | |
| QY | 1 | MAWKTLPIYLLLLSVFVIQQVSSQ----- | 25 | | | | |
| Db | 1 | MAWKTLPIYLLLLSVFVIQQVSSQDLSSCAGRCGEGYSRDATCNDCYNQHYMECCPDF | 60 | | | | |
| QY | 26 | -----ELSKGRCFESFERGECDCDAQCKKYDKCCPDYESFCAEHSVSENQSSSS | 79 | | | | |
| Db | 61 | KRVCTAELSKGRCFESFERGECDCDAQCKKYDKCCPDYESFCAE----- | 106 | | | | |
| QY | 80 | SSSSSSSTTWIKSSKNSAANRELQKLVKNQKNRKKKTPKPPVVDAGSLDNGD | 139 | | | | |
| Db | 107 | -----VKDNKNRKKKTPKPPVVDAGSLDNGD | 137 | | | | |
| QY | 140 | FKVTTPTDSTTQHNKYSTSPKITTAKPINPRSLPNSDTSKETSILTNNKETTIVTKETT | 199 | | | | |
| Db | 138 | FKVTTPTDSTTQHNKYSTSPKITTAKPINPRSLPNSDTSKETSILTNNKETTIVTKETT | 197 | | | | |


```
QY 200 TTNKQSTGKSKTSAKETQSIKTSAXDLAPTQSVLAKPTPKAETTTKGALATPKPEP 259
DB 198 TTNKQSTGKSKTSAKETQSIKTSAXDLAPTQSVLAKPTPKAETTTKGALATPKPEP 257
QY 260 TPTTPKEPASTTPKEPTPTTIKSAPTTPKEPAPTTTKSAPTTTPKEPAPTTTKKEPAPTTTPK 319
DB 258 TPTTPKEPASTTPKEPTPTTIKSAPTTPKEPAPTTTKSAPTTTPKEPAPTTTKKEPAPTTTPK 317
QY 320 EPAPTTTKPAPTTTKSAPTTTPKEPAPTTPKAPATTPPKAPATTPPKAPATTPPKAPATTP 379
DB 318 EPAPTTTKPAPTTTKSAPTTTPKEPAPTTPKAPATTPPKAPATTPPKAPATTPPKAPATTP 377
QY 380 TKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKSA 439
DB 378 TKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKSA 437
QY 440 PTTTKPAPTTTKSAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTT 499
DB 438 PTTTKPAPTTTKSAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTT 497
QY 500 EPAPTTTKPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTPEELAP 559
DB 498 EPAPTTTKPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTPEELAP 557
QY 560 TTPPEEPTPTTPKEPAPTTTPKAAAPNTPKAPATTPPKAPATTPPKAPATTPPKETAAPTTPK 619
DB 558 TTPPEEPTPTTPKEPAPTTTPKAAAPNTPKAPATTPPKAPATTPPKAPATTPPKETAAPTTPK 617
QY 620 GTAPTTLKAPATTPPKAPKELAPTTTKEPTSTSDKAPATTPKGTAPTTKEPAPTTTP 679
DB 618 GTAPTTLKAPATTPPKAPKELAPTTTKEPTSTSDKAPATTPKGTAPTTKEPAPTTTP 677
QY 680 KEPAPTTTPKGTAPTTLKAPATTPPKAPKELAPTTTKEPTSTSDKAPATTPKGTAPTT 739
DB 678 KEPAPTTTPKGTAPTTLKAPATTPPKAPKELAPTTTKEPTSTSDKAPATTPKGTAPTT 737
QY 740 PKPAPATTPPKAPATTPPTTPPTTSEVSTPTTKGPTTIHKSPDSESTPELSAETPKALE 799
DB 738 PKPAPATTPPKAPATTPPTTPPTTSEVSTPTTKGPTTIHKSPDSESTPELSAETPKALE 797
QY 800 NSPKPVGPTTKPAATKPEMTTAKDKUTERDLRTTPTTTAAAGMTKETATTTTEKTE 859
DB 798 NSPKPVGPTTKPAATKPEMTTAKDKUTERDLRTTPTTTAAAGMTKETATTTTEKTE 857
QY 860 SKITATTTQVSTTTQDTTPFKITLLKTTTLAPKVTTTKKTTITTTTEIMNKPEETAKPKDR 919
DB 858 SKITATTTQVSTTTQDTTPFKITLLKTTTLAPKVTTTKKTTITTTTEIMNKPEETAKPKDR 917
QY 920 ATNSKATTPKPKQPTKAPKPTSTKPKTMPRVRKPTTPTRKMTSTWPELNPTSRITAE 979
DB 918 ATNSKATTPKPKQPTKAPKPTSTKPKTMPRVRKPTTPTRKMTSTWPELNPTSRITAE 977
QY 980 AMLQTTTRNQTNSKLVNPKSEDAAGAGSETPHMLLRPHVFWPEVTPDMDYLPRVFN 1039
DB 978 AMLQTTTRNQTNSKLVNPKSEDAAGAGSETPHMLLRPHVFWPEVTPDMDYLPRVFN 1037
QY 1040 QGIIINPMLS 1049
DB 1038 QGIIINPMLS 1047
```

RESULT 11

ABU53252

ID ABU53252 standard; protein; 546 AA.

XX

AC ABU53252;

XX

DT 14-APR-2003 (first entry)

XX

DE Human testes-derived DKF2phtes3_4019 homologue #1.

XX

KW Human; gene therapy; vaccine; disease treatment; detection.

XX

OS Homo sapiens.
XX WO200112659-A2.
XX
XX 22-FEB-2001.
XX
XX 18-AUG-2000; 2000WO-IB001496.
XX
XX 18-AUG-1999; 99US-0149499P.
XX 28-SEP-1999; 99US-0156503P.
XX (GBHU-) GERMAN HUMAN GENOME PROJECT.
XX
XX Wiemann S;
XX
XX WPI; 2001-327840/34.
XX
XX Nucleic acids having the sequences of clones isolated from libraries of
XX different human tissues, useful in recombinant DNA methodologies.
XX
XX Example III; Page 892; 1095pp; English.
XX
XX This invention describes novel polynucleotides and polypeptides isolated
XX from human cDNA libraries which can be used for gene therapy or in
XX vaccines. The polynucleotides of the invention and antibodies encoded by
XX them may be used in the prevention, diagnosis and treatment of diseases
XX associated with inappropriate polypeptide expression. The products of the
XX invention may also be used to identify modulators of expression and
XX activity and to down regulate expression and activity. The antibodies of
XX the invention may also be used as diagnostic agents for detecting the
XX presence of polypeptides in samples. This sequence represents a homologue
XX of a polypeptide described in the disclosure of the invention
XX
XX Sequence 546 AA;

Query March 52.4%; Score 2920; DB 4; Length 546;

Best Local Similarity 100.0%; Pred. NO. 1.5e-16;

Matches 546; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
QY 397 KKPAPTTPKEPAPTTTPKEPAPTTTKESPTTPKEPAPTTTKSAPTTTKAPATTTKSAPT 456
DB 1 KKPAPTTPKEPAPTTTPKEPAPTTTKESPTTPKEPAPTTTKSAPTTTKAPATTTKSAPT 60
QY 457 TPKEPSTTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTKAPATPK 516
DB 61 TPKEPSTTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTKAPATPK 120
QY 517 PAPTPKETAPTTPEKLTPTTPEKLAAPTPEKAPATTPPELAPTTPEEPPTTPEEPAPT 576
DB 121 PAPTPKETAPTTPEKLTPTTPEKLAAPTPEKAPATTPPELAPTTPEEPPTTPEEPAPT 180
QY 577 TPKAAAPNTPKPAPTTTPKEPAPTTTPKEPAPTTTPKETAPTTTPKETAPTTTLKEPAPTTPKK 636
DB 181 TPKAAAPNTPKPAPTTTPKEPAPTTTPKEPAPTTTPKETAPTTTPKETAPTTTLKEPAPTTPKK 240
QY 637 PAPKELAPTTTKEPTSTSDKAPATTPKGTAPTTTPKEPAPTTTPKEPAPTTTPKGTAPTLK 696
DB 241 PAPKELAPTTTKEPTSTSDKAPATTPKGTAPTTTPKEPAPTTTPKEPAPTTTPKGTAPTLK 300
QY 697 EPAPTTTPKPKAPKELAPTTTKGPTSTSDKAPATTPPKETAPTTTPKEPAPTTTPKAPATTP 756
DB 301 EPAPTTTPKPKAPKELAPTTTKGPTSTSDKAPATTPPKETAPTTTPKEPAPTTTPKAPATTP 360
QY 757 ETDPPTTSEVSTPTTTPKEPTTIHKSPDSESTPELSAETPKALENSPKPVGPTTKTAAAT 816
DB 361 ETDPPTTSEVSTPTTTPKEPTTIHKSPDSESTPELSAETPKALENSPKPVGPTTKTAAAT 420
QY 817 KPEMTTAKDKUTERDLRTTPTTTAAAGMTKETATTTTEKTTESKITATTTQVSTTTTQD 876
DB 421 KPEMTTAKDKUTERDLRTTPTTTAAAGMTKETATTTTEKTTESKITATTTQVSTTTTQD 480
QY 877 TTPFKITLLKTTTLAPKVTTTKKTTITTTTEIMNKPEETAKPKDRATNSKATTPKPKQPTKA 936
DB
```


